SHORT REPORT

Diagnosis by sequencing: correction of misdiagnosis from FSHD2 to LGMD2A by whole-exome analysis

Andreas Leidenroth¹, Hanne Sørmo Sorte², Gregor Gilfillan², Melanie Ehrlich³, Robert Lyle² and Jane E Hewitt^{*,1}

We studied and validated facioscapulohumeral muscular dystrophy (FSHD) samples from patients without a D4Z4 contraction (FSHD2 or 'phenotypic FSHD'). For this, we developed non-radioactive protocols to test D4Z4 allele constitution and DNA methylation, and applied these to samples from the Coriell Institute Cell Repository. The D4Z4 sizing showed two related subjects to have classic chromosome 4 contraction-dependent FSHD1. A third sample (GM17726) did not have a short chromosome 4 fragment, and had been assigned as non-4q FSHD (FSHD2). We tested D4Z4 haplotype and methylation for this individual but found both to be inconsistent with this diagnosis. Using exome sequencing, we identified two known pathogenic mutations in *CAPN3* (Arg490GIn and Thr184Argfs*36), indicating a case of LGMD2A rather than FSHD. Our study shows how a wrong diagnosis can easily be corrected by whole-exome sequencing by constraining the variant analysis to candidate genes after the data have been generated. This new way of 'diagnosis by sequencing' is likely to become common place in genetic diagnostic laboratories. We also publish a digoxigenin-labeled Southern protocol to test D4Z4 methylation. Our data supports hypomethylation as a good epigenetic predictor for FSHD2. The non-radioactive protocol will help to make this assay more accessible to clinical diagnostic laboratories and the wider FSHD research community.

European Journal of Human Genetics (2012) 20, 999–1003; doi:10.1038/ejhg.2012.42; published online 29 February 2012

Keywords: exome sequencing; methylation; facioscapulohumeral muscular dystrophy; limb-girdle muscular dystrophy

INTRODUCTION

Facioscapulohumeral muscular dystrophy (FSHD) is a progressive disorder that varies in clinical severity from wheel chair dependency to minor impairments. The most common form of the disease (FSHD1) is associated with contractions of the tandem repeat array D4Z4 in the subtelomeric region of chromosome 4.¹ D4Z4 is composed of many copies of a 3.3-kb repeat unit arranged head-to-tail. Healthy individuals usually have 11–150 repeats, whereas FSHD1 patients have one 4q allele with fewer than 11 repeats. To test for FSHD, the D4Z4 array is sized by digestion of genomic DNA with the restriction enzyme *Eco*RI, gel electrophoresis and Southern blotting with the p13E-11 probe (Figure 1a). In over 90% of cases, this identifies a contracted allele.^{1,2}

Only chromosomes of a particular haplotype are 'disease-permissive', which explains why not all D4Z4 contractions are pathogenic. Lemmers *et al*^{3,4} genotyped single-nucleotide polymorphisms (SNPs) in the p13E-11 region, a D4Z4-proximal short sequence length polymorphism (SSLP) and a structural polymorphism (A- or B-type) in the D4Z4-distal pLAM region in 300 patients and controls. These studies showed that FSHD develops only if D4Z4 is deleted on a disease-permissive haplotype known as '4qA161', or on closely related rarer haplotypes. 4qA161 is defined by a 161-bp SSLP, specific p13E-11 SNPs and an 'A'-type pLAM structural variant (Figure 1a). The important difference between 4qA161 and non-permissive haplotypes lies at the distal end of D4Z4. On the 4qA161 haplotype, pLAM contains a functional polyadenylation signal that is in linkage disequilibrium with the p13E-11 and SSLP variants.^{3,5,6} In FSHD, polyadenylation stabilizes aberrant D4Z4 transcripts that are thought to be disease-causative.^{5,7} On non-permissive haplotypes, this signal is absent.

Pathogenic splice-forms of the transcripts are only generated when the local chromatin is de-repressed. Repeat contraction reduces the DNA methylation and histone 3 lysine 9 tri-methylation (H3K9me3) that normally repress D4Z4 chromatin.² This increases transcription, thereby generating specific RNA isoforms of the *DUX4* copy in the last repeat that are polyadenylated on 4qA161 and cause disease.^{5,7}

About 10% of patients clinically diagnosed with FSHD have no D4Z4 contraction ('phenotypic FSHD' or FSHD2).⁸ However, they share the loss of DNA methylation and H3K9me3 and the requirement for a permissive 4q haplotype with FSHD1 patients.⁹⁻¹¹ Although FSHD2 often occurs sporadically, a number of familial cases have also been reported, which suggests an inherited component.⁸ It is possible that FSHD2 patients lose methylation due to a mutation at a genetic locus different from D4Z4, but that they share the same *DUX4*-dependent downstream disease pathway and phenotype. Thus, D4Z4 methylation is an epigenetic marker that unites FSHD1 and FSHD2 and provides a diagnostic test to distinguish them both from unrelated muscular dystrophies. Because the mutation(s) that cause FSHD2 are still unknown, we are particularly interested in this subset of patients.

¹Centre for Genetics and Genomics, School of Biology, Queen's Medical Centre, The University of Nottingham, Nottingham, UK; ²Department of Medical Genetics, Oslo University Hospital, Oslo, Norway; ³Human Genetics Program and Tulane Cancer Center, Tulane Medical School, New Orleans, LA, USA

^{*}Correspondence: Dr JE Hewitt, Centre for Genetics and Genomics, School of Biology, Queen's Medical Centre, The University of Nottingham, Nottingham NG7 2UH, UK. Tel: +44 (0) 115 823 0364; Fax: +44 (0) 115 823 0338; E-mail: jane.hewitt@nottingham.ac.uk

Received 14 December 2011; revised 23 January 2012; accepted 31 January 2012; published online 29 February 2012



Figure 1 Genetic and epigenetic characterization of GM17726. Not to scale. (a) Locations of probes shaded in gray. Upper panel shows restriction enzyme sites used in pulsed-field analysis. Lower panel shows methylation assay enzymes and predicted fragment sizes. Adapted from de Greef *et al.*⁸ X=Xapl. (b) The four D4Z4 alleles of subject GM17726 were separated by PFGE. *Bln*I normally cuts only within chromosome 10-derived repeats, whereas *Xapl* cuts only within chromosome 4-derived repeats. Therefore, the two smaller *Xapl*-resistant alleles are from chromosome 10, whereas the larger *Bln*I-resistant alleles are from chromosome 4. Both *Bln*I and *Xapl* cut closer to the D4Z4 repeats than *EcoR*I, which slightly decreases the total fragment size. (c) GM17726 has three A-type telomeres. The probe also hybridizes to the 10qA telomere. For hybridization with the A-probe, genomic DNA was digested with *Hind*III instead of *EcoR*I. The blot was first hybridized with p13E-11, then stripped and re-hybridized with the A-probe. (d) The SSLP PCR is sized by capillary electrophoresis with known standards. GM17726 carries one 163 bp and three 166 bp alleles, but no 161 bp allele. (e) Restriction digests with *Bg/I* combined with the methylation-sensitive enzymes *Fsel* and *BsaA*I show the hypomethylated chromosome 4 alleles in FSHD siblings and the ICF control. *Bln*I is included in the digest to eliminate chromosome 10-derived alleles. GM17726 has high methylation.

METHODS

Detailed methods are included in the supplement.

Sample acquisition

Samples were purchased as lymphoblast cell lines from the Coriell Cell Repository. Family 1948 included two affected individuals (GM16348 and GM16351) and their unaffected sister (GM16352). GM16348 and GM16351 have both been deposited as 'FSHD type unknown' with a clinical diagnosis consistent with FSHD but no genetic information. Subject GM17726 presented with scapular winging, facial muscle weakness, progressive bilateral weakness of the arms and legs, and elevated CPK, which was clinically diagnosed as FSHD. However, the D4Z4 sizes are reported as 28+59 repeats on chromosome 4 and 5+22 repeats on chromosome 10. Accordingly, the Coriell catalog lists the diagnosis as 'non-4q FSHD'.

Pulsed-field gel electrophoresis and non-radioactive probes

D4Z4 allele sizes, subtelomeric polymorphisms and methylation were tested by pulsed-field gel electrophoresis and Southern blotting as published,^{8,12} but with non-radioactive digoxigenin-labeled probes instead of conventional ³²P labeling. The p13E-11 probe was amplified (primers by Kekou *et al*¹³) using Bioline BioMix Red (Bioline, London, UK) and cloned into pGEM T-Easy. Probe was synthesized from 30 pg of plasmid with the Roche PCR DIG kit (Roche, London, UK). We further designed a probe to the A-type telomere using primers 5'-TGGGAATACTGAACACAGAAATGA-3' and 5'-GCATGACAATTTC AGACTCCA-3', which also hybridizes to the closely related 10qA telomere. 50 pg of construct LS1093-1¹⁴ was used as template for the A-probe. PCRs were performed in 50µl reactions with a 1:1 mix of Roche dNTP mix and DIG-labeled dUTP mix and the following programme: 95 °C for 10 min (hot-start), 10 cycles of 95° C for 30 s, 60° C (p13E-11) or 54.8° C (A-probe)

for 30 s, 68° C for 30 s followed by 32 cycles of 98° C for 30 s, 60° or 54.8° C for 30 s, and 68° C for 45 s with an additional 20 s increase/cycle, with a final step at 68° C for 15 min. Non-radioactive Southern blot analysis was performed as described in the supplement.

SSLP genotyping

SSLPs were typed as published,³ but with a FAM-labeled primer (Sigma-Aldrich, St Louis, MO, USA).

Exome capture and sequencing

Exome capture was performed with the Agilent SureSelect Target Enrichment System (Agilent, Santa Clara, CA, USA) for the Illumina Paired-End Multiplexed Sequencing Library 1.2 (Illumina, San Diego, CA, USA). Data were processed on an iMac with open-source tools using a custom Perl pipeline.^{15–18} After the initial mapping with BWA,¹⁸ reads were re-aligned around sites of known insertions and deletions with the Genome Analysis Toolkit (GATK).¹⁷ Next, likely PCR duplicates were removed with SAMtools and Picard. Finally, the base-quality scores were recalibrated using GATK. Variants were called using the GATK UnifiedGenotyper, and low confidence variant calls were filtered with the GATK FiltrationWalker.¹⁶ Calls were annotated (http://snp.gs.washington.edu/SeattleSeqAnnotation/) and analyzed with Perl scripts and Excel.

RT-PCR

GM17726 RNA was extracted with TRIzol (Invitrogen, Carlsbad, CA, USA) and $10 \,\mu g$ of RNA was DNAase treated with the Ambion TURBO DNA-free kit (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions.







1002

cDNA was synthesized from $2 \mu g$ RNA using the Superscript III system (Invitrogen) with 0.5 μg random primers (Invitrogen) in first-strand synthesis buffer according to the Superscript III manual. *CAPN3* exons 3–13 were amplified with BioMix Red (Bioline) in a single product of 1098 bp, cloned into pGEM T-Easy (Promega, Fitchburg, WI, USA) and sequenced.

RESULTS

Both subjects from family 1948 share a short chromosome 4 allele of about 18 kb, which is not present in their healthy sister (Supplementary Figure S1A). For GM17726 we confirmed the reported allele sizes (Figure 1b). The subject carries three A-type alleles and, by exclusion, one B-type allele (Figure 1c). When we genotyped the SSLP,³ we found that this individual carries one 163 bp and three 166 bp alleles, but no 161-bp allele (Figure 1d).

When we tested the proximal D4Z4 unit for DNA methylation, we detected high methylation levels for GM17726, which were comparable to those we observed in unaffected controls (Figure 1e). As a positive control, we used DNA from an ICF syndrome lymphoblast cell line (GM08714). This syndrome is caused by mutations in the DNA methyltransferase *DNMT3b*, and patients show almost complete loss of D4Z4 methylation (Figure 1e).^{8,19} The unaffected individual (GM16352) from family 1948 shows high methylation levels, while her affected siblings GM16351 and GM16348 are hypomethylated. Instead of FSHD2, family 1948 likely represents a standard case of contraction-dependent FSHD1, and the methylation data are consistent with this. We therefore did not study these samples any further, but will update the catalog.

On the basis of the published haplotype frequencies,⁶ the most likely allele constitution of GM17726 is 1× 4qB163, 1× 4qA166 and 2×10 qA166. FSHD depends on the presence of at least one permissive allele; 4gB163 and 4gA166 have both been reported as being nonpermissive.^{3–6} Although one recent study identified a small number of patients with apparently causative short D4Z4 arrays on these nonpermissive haplotypes, DUX4 expression was not confirmed in these individuals.²⁰ Furthermore, as subject GM17726 also had no chromosome 4 contraction and normal methylation levels, we predicted a different muscular dystrophy and investigated this using exome sequencing. This produced 77.2 million paired-end read pairs. Eighty-seven percent of single reads mapped uniquely to the GRCh37 reference assembly. Of these, 77% mapped to the capture targets, an average read-depth of 172 reads/base. After the quality processing, the average depth across all targets was 59 reads/ base, with 71% on target. 85% had a mapping quality score of 30 or higher.

High-confidence variants were annotated using SeattleSeq (http://snp.gs.washington.edu/SeattleSeqAnnotation/). We detected 33499 single-nucleotide variants, of which 32538 had been previously reported in dbSNP131 or the 1000 Genome Project. Excluding these left 961 variants, of which 11 were nonsense and 424 missense. Using the GATK IndelGenotyperV2.0, we also identified 51 frameshift insertion/deletion (indel) mutations. We intersected the gene list of the 435 coding variants and 51 frameshift indels with 48 genes associated with different muscular dystrophies (www.dmd.nl). This identified two different variants in CAPN3, which had been previously reported in limb-girdle muscular dystrophy type 2A (LGMD2A). The mutations were confirmed by Sanger sequencing (Figures 2a and b). In exon 11, a G>A change causes an Arg490Gln substitution. In exon 4, deletion of a single adenine results in a Thr184Argfs*36 frame shift. RT-PCR and sequencing confirmed that these two mutations are on different chromosomes (Figure 2c).

DISCUSSION

Our data support the view that the 'FSHD2' label should only be applied if the patient has been tested for D4Z4 hypomethylation and carries at least one chromosome of permissive haplotype. Unless all of these criteria are met, LGMD2A should always be considered as a differential diagnosis if no short chromosome 4qA fragment is found. This agrees with the results of a recent candidate gene screen of patients who had been initially diagnosed with FSHD, but lacked D4Z4 deletions.²¹

Furthermore, we assigned the genetic diagnosis in an isolated case using exome sequencing. This was possible without a family pedigree or direct access to the patient. A recent report by Lim *et al*²² demonstrated the successful use of target-enrichment sequencing to identify mutations in Duchenne and Becker's muscular dystrophy. The authors designed a custom Agilent capture library targeting dystrophin and 25 other muscular dystrophy genes. Our approach was analogous, but in contrast we only restricted our candidate gene list during data analysis rather than data generation. The advantage of this approach is that if no causative mutation is identified after an initial analysis, the search could easily be extended to the whole exome. This avoids the cost and time delay of additional laboratory analyses.

This 'diagnosis by sequencing' that we have demonstrated here is likely to become a standard tool in clinical genetics laboratories around the world. In addition, the non-radioactive protocol modifications we made to the D4Z4 methylation assay will make it easier for clinical diagnostic laboratories and researchers to test for D4Z4 methylation in FSHD2.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Tim Smith and Jess Tyson for their technical assistance. AL is funded by a PhD studentship from the Muscular Dystrophy Campaign UK. A University of Nottingham BESTS travel prize is gratefully acknowledged.

- Wijmenga C, Hewitt JE, Sandkuijl LA *et al*: Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* 1992; 2: 26–30.
 van der Maarel SM, Tawil R, Tapscott SJ: Facioscapulohumeral muscular dystrophy and
- 2 Van der Maarel SM, Tawil R, Tapscott SJ: Facioscapulonumeral muscular dystropny and DUX4: breaking the silence. *Trends Mol Med* 2011; **17**: 252–258.
- 3 Lemmers R, 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.
- 4 Lemmers R, 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.
- 5 Lemmers R, van der Vliet PJ, Klooster R et al: A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 2010; **329**: 1650–1653.
- 6 Lemmers R, van der Vliet PJ, van der Gaag KJ et al: Worldwide population analysis of the 4q and 10q subtelomeres identifies only four discrete interchromosomal sequence transfers in human evolution. Am J Hum Genet 2010; 86: 364–377.
- 7 Snider L, Geng LN, Lemmers RJLF et al: Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS Genet 2010; 6: e1001181.
- 8 de Greef JC, Lemmers RJL, van Engelen BGM *et al*: Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum Mutat* 2009; **30**: 1449–1459.
- 9 van Overveld PGM, Lemmers R, Sandkuijl LA *et al*: Hypomethylation of D4Z4 in 4qlinked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet* 2003; 35: 315–317.
- 10 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.
- 11 Zeng W, de Greef JC, Chen Y-Y et al: Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS Genet 2009; 5: e1000559.
- 12 Lemmers RJL, de Kievit P, van Geel M *et al*: Complete allele information in the diagnosis of facioscapulohumeral muscular dystrophy by triple DNA analysis. *Ann Neurol* 2001; **50**: 816–819.

- 13 Kekou K, Fryssira H, Sophocleous C, Mavrou A, Manta P, Metaxotou C: Facioscapulohumeral muscular dystrophy molecular testing using a non radioactive protocol. *Mol Cell Probes* 2005; **19**: 422–424.
- 14 van Geel M, Dickson MC, Beck AF *et al*: Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. *Genomics* 2002; **79**: 210–217.
- 15 Li H, Handsaker B, Wysoker A et al: The sequence alignment/map format and SAMtools. Bioinformatics 2009; 25: 2078–2079.
- 16 DePristo MA, Banks E, Poplin R *et al*: A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; 43: 491–498.
- 17 McKenna A, Hanna M, Banks E et al: The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20: 1297–1303.
- 18 Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010; 26: 589–595.

- 19 Hansen RS, Wijmenga C, Luo P et al: The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A 1999; 96: 14412–14417.
- 20 Scionti I, Fabbri G, Fiorillo C et al: Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. J Med Genet 2012; e-pub ahead of print 3 January 2012; doi:10.1136/ jmedgenet-2011-100454.
- 21 Sacconi S, Camano P, de Greef JC *et al*: Patients with a phenotype consistent with facioscapulohumeral muscular dystrophy display genetic and epigenetic heterogeneity. *J Med Genet* 2012; **49**: 41–46.
- 22 Lim BC, Lee S, Shin JY et al: Genetic diagnosis of Duchenne and Becker muscular dystrophy using next-generation sequencing technology: comprehensive mutational search in a single platform. J Med Genet 2011; 48: 731–736.
- 23 Robinson JT, Thorvaldsdottir H, Winckler W et al: Integrative genomics viewer. Nat Biotechnol 2011; 29: 24–26.

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