


FSHD muscle shows perturbation in fibroadipogenic progenitor cells, mitochondrial function and alternative splicing independently of inflammation

Elise N. Engquist¹, Anna Greco^{2,3}, Leo A.B. Joosten^{3,4}, Baziel G.M. van Engelen^{2,†}, Peter S. Zammit ^{1,†,*}, Christopher R.S. Banerji^{1,5,†}

¹Randall Centre for Cell and Molecular Biophysics, King's College London, New Hunt's House, Guy's Campus, London SE1 1UL, United Kingdom

²Department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, 6525 GA, The Netherlands

³Department of Internal Medicine, Radboud Institute of Molecular Life Sciences (RIMLS) and Radboud Center of Infectious Diseases (RCI), Radboud University Medical Center, Geert Groteplein Zuid 10, Nijmegen 6525 GA, The Netherlands

⁴Department of Medical Genetics, Iuliu Hatieganu University of Medicine and Pharmacy, 400012, Cluj-Napoca, Romania

⁵The Alan Turing Institute, The British Library, 96 Euston Road, London NW1 2DB, United Kingdom

*Corresponding author. Peter S. Zammit, King's College London, Randall Centre for Cell and Molecular Biophysics, New Hunt's House, Guy's Campus, London SE1 1UL, United Kingdom. E-mail: peter.zammit@kcl.ac.uk

†Baziel G. M. van Engelen, Peter S. Zammit and Christopher R. S. Banerji are joint senior authors.

Abstract

Facioscapulohumeral muscular dystrophy (FSHD) is a prevalent, incurable myopathy. FSHD is highly heterogeneous, with patients following a variety of clinical trajectories, complicating clinical trials. Skeletal muscle in FSHD undergoes fibrosis and fatty replacement that can be accelerated by inflammation, adding to heterogeneity. Well controlled molecular studies are thus essential to both categorize FSHD patients into distinct subtypes and understand pathomechanisms. Here, we further analyzed RNA-sequencing data from 24 FSHD patients, each of whom donated a biopsy from both a non-inflamed (TIRM⁻) and inflamed (TIRM⁺) muscle, and 15 FSHD patients who donated peripheral blood mononucleated cells (PBMCs), alongside non-affected control individuals. Differential gene expression analysis identified suppression of mitochondrial biogenesis and up-regulation of fibroadipogenic progenitor (FAP) gene expression in FSHD muscle, which was particularly marked on inflamed samples. PBMCs demonstrated suppression of antigen presentation in FSHD. Gene expression deconvolution revealed FAP expansion as a consistent feature of FSHD muscle, via meta-analysis of 7 independent transcriptomic datasets. Clustering of muscle biopsies separated patients in an unbiased manner into clinically mild and severe subtypes, independently of known disease modifiers (age, sex, D4Z4 repeat length). Lastly, the first genome-wide analysis of alternative splicing in FSHD muscle revealed perturbation of autophagy, BMP2 and HMGB1 signalling. Overall, our findings reveal molecular subtypes of FSHD with clinical relevance and identify novel pathomechanisms for this highly heterogeneous condition.

Keywords: facioscapulohumeral muscular dystrophy (FSHD); fibroadipogenic progenitor cells (FAPs); transcriptomics; mitochondrial function; alternative splicing

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) is an autosomal dominant myopathy with a prevalence of ~12/100 000 [1] that typically presents during the second or third decade of life. FSHD is characterized by muscle weakness and a progressive loss of muscle mass that typically begins in the facial muscles before progressing to the shoulder girdle and lower limb [2–4]. However, a high degree of inter-patient variability is observed in age of onset, muscle groups affected and rate of progression. FSHD also associates with extra-muscular features including sensorineural, high-frequency hearing loss and retinal telangiectasia, with increased prevalence in more severely affected patients [5–7].

FSHD comprises two genetically distinct subtypes: FSHD1 (OMIM: 158900, ~95% of cases) and FSHD2 (OMIM: 158901, ~5% of cases). Both subtypes result in epigenetic derepression of the D4Z4 macrosatellite region in the telomeric region of chromosome 4q35 [8]. This is due to contraction of the D4Z4 region from the healthy range of 100–11 D4Z4 repeats to 10–1 on at least one allele in

FSHD1 [9, 10], or mutations in chromatin modifiers in FSHD2 [11, 12]. Either event enables transcription of the *Double homeobox 4* (*DUX4*) (OMIM: 606009) retrogene from the distal-most D4Z4 unit [13–15]. When combined in cis with a permissive 4qA haplotype supplying a poly(A) signal, the stabilized ectopic *DUX4* mRNA [11] can be translated to generate *DUX4* protein, which is believed to be the key driver of FSHD pathogenesis [7, 16, 17].

DUX4 is a pioneer transcription factor involved in zygotic genome activation during the cleavage stage [18], which is then epigenetically silenced in somatic tissues. Re-expression of *DUX4* and its target genes in FSHD is linked to myotoxicity both *in vitro* [3, 19–26] and *in vivo* [27–31]. Mechanisms of *DUX4* myotoxicity include mitochondrial dysfunction [21, 32–34], altered muscle regeneration through functional interference with *PAX7* [26, 35, 36], increased exposure and susceptibility to oxidative stress [22, 37–40], p53-dependent apoptosis [24, 25] and alterations in RNA splicing [41–43]. However, full understanding of the mechanisms through which *DUX4* drives pathology in FSHD remains elusive.

Received: April 13, 2023. Revised: September 25, 2023. Accepted: October 10, 2023

© The Author(s) 2023. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Although DUX4 is very difficult to detect/undetectable in patient biopsies using routine lab methods [15], signatures based on DUX4 target gene expression, or PAX7 target gene repression, associate with pathological severity and disease duration [26, 35, 36]. Further pathomechanisms identified from discovery-based analysis of FSHD muscle biopsy transcriptomics include Wnt/ β -catenin signalling [44, 45], HIF1 α overactivation [26, 44], TNF α overactivation [21, 44], mis-regulation of vascular genes [46], perturbed myogenesis [47], altered inflammatory signalling [48, 49] and RAGE-NF- κ b signalling [39].

FSHD muscle is characterized by inflammation, which accelerates underlying fatty replacement of muscle and drives clinical weakness [7, 50]. As such, histological and transcriptomic analyses of skeletal muscle of FSHD patients are often accompanied by magnetic resonance imaging (MRI) to select inflamed muscles, which are identifiable by oedema using inversion recovery MRI sequences (e.g. TIRM, STIR) [51–55]. Such studies demonstrate that numbers of regenerating fibres in FSHD muscle positively correlate with severity [54] and inform several transcriptomic signatures of FSHD that can distinguish FSHD muscle from control with varying degrees of accuracy [26, 36, 49, 55–57].

Myogenic cells are not the only perturbed cell type in FSHD, implicating non-myogenic cell types resident in muscle, particularly fibroadipogenic progenitor (FAP) cells [53, 58]. FAPs comprise a heterogeneous mesenchymal cell population, which interact with immune cells and muscle stem cells to facilitate muscle regeneration. Mis-regulation of FAPs contributes to fibrosis and fat infiltration in other forms of muscular dystrophy [59]. Recent work indicates a similar mis-regulation of FAPs may exist in FSHD [53, 58]. Peripheral blood mononuclear cells (PBMCs) from FSHD patients have been examined to develop minimally invasive blood-based biomarkers of disease progression. These studies suggest a potential role for IL-6 and identify transcriptomic signatures shared across FSHD muscle and blood [48, 60–62].

In addition to genetics, there are many sources of heterogeneity in FSHD muscle, with factors such as inflammation driving intra-patient variability, and external factors such as lifestyle, contributing to inter-patient variability. As such, clinical heterogeneity in FSHD patients must be controlled when evaluating tissue-specific effects such as inflammation. Well-designed transcriptomic studies of multiple FSHD tissues are essential to better understand molecular features of FSHD, facilitate biomarker discovery and guide therapeutic development.

Here, we report further transcriptomic analysis of isogenic MRI-guided biopsies of both non-inflamed (TIRM⁻) and actively inflamed (TIRM⁺) muscles and PBMCs from the same FSHD patients, alongside muscle-matched biopsies and PBMCs from healthy controls [60]. Our analyses indicate FAP population expansion and mitochondrial dysfunction in FSHD muscle, which progress with inflammation. Our transcript-level investigation provides global analysis of RNA splicing in FSHD muscle, detecting differences in post-transcriptional regulation with pathological relevance that were previously masked by gene-level aggregation. Finally, we describe potential defects in antigen presentation in FSHD PBMCs.

Results

Differential gene expression of FSHD muscle biopsies reveals transcriptomic profiles associated with clinical severity

We further considered our data corresponding to 49 individuals: 25 FSHD1 patients, 1 FSHD2 patient and 23 control

individuals, described in detail in [60]. Briefly, clinical variables obtained included: age, sex and MRC sum score for 12 muscle groups. For FSHD patients, severity indicators measured included: D4Z4 repeat length (for FSHD1 patients), Ricci clinical severity score [63], Lamperti clinical severity score [64], disease duration, lower limb fat fraction (LLFF) assessed by MRI and maximum voluntary contraction (MVC) of tibialis anterior (TA) [60].

As we described previously [60], muscle biopsies were obtained from 35/49 individuals for RNA-sequencing. 23 FSHD1 patients and 1 FSHD2 patient underwent MRI guided muscle biopsies with samples taken from TIRM⁻ (23/24 vastus lateralis) and TIRM⁺ muscle (15/24 gastrocnemius) from each patient. 11 control individuals also donated a muscle biopsy of vastus lateralis. FSHD patients who underwent muscle biopsy were significantly older than controls and showed a male sex bias [60].

Two separate differential gene expression analyses were performed on the muscle biopsy data, analysing length adjusted transcript counts aggregated to gene level via negative binomial models using the DESeq2 package [65]. First, we performed a multivariate analysis to identify genes whose expression differed across control, TIRM⁻ and TIRM⁺ samples, adjusting for age and sex. Comparing TIRM⁻ FSHD samples to controls independently of TIRM⁺, we identified 17 genes down-regulated and 40 genes up-regulated after adjustment for multiple testing. Comparing TIRM⁺ FSHD samples to control independent of TIRM⁻, we identified 1650 genes up-regulated and 542 down-regulated after adjustment for multiple testing (Fig. 1A and B, Supplementary Table S1). This suggests that TIRM⁻ samples are more subtly changed from control muscle in terms of the gene expression landscape compared to TIRM⁺ samples.

Second, we focused on FSHD samples and performed a paired analysis comparing gene expression in isogenic TIRM⁻ and TIRM⁺ samples from each FSHD patient. Directly comparing isogenic TIRM⁻ and TIRM⁺ FSHD samples, we identified 2127 genes up-regulated in TIRM⁺ samples compared to matched TIRM⁻ samples and 1484 genes down-regulated. There were overlaps in genes identified in the two analyses, with 16 genes consistently up-regulated in TIRM⁻ versus control, TIRM⁺ versus control and TIRM⁺ versus TIRM⁻, but no genes consistently down-regulated in the three comparisons (Fig. 1A and B).

DUX4 was only detected in 2 TIRM⁻ and 4 non-corresponding TIRM⁺ FSHD samples, and undetectable in both the remaining 42/48 FSHD samples and all (11/11) control samples. Where detected, DUX4 was at very low levels (max 3 reads per sample) and was not differentially expressed between FSHD and control samples, nor between matched TIRM⁻ and TIRM⁺ samples.

We next performed a clustering analysis of FSHD and control samples. As a feature space, we used all genes differentially expressed in TIRM⁻ samples versus controls and the 500 most significantly up-regulated or down-regulated genes in TIRM⁺ samples versus controls, as assessed by adjusted P-value (1015 genes in total). K-medoids clustering using a Pearson correlation distance metric was performed using the ConsensusClusterPlus algorithm [66] and cluster stability plots indicated an 11 cluster solution (Fig. 1C). Principal component analysis (PCA) demonstrated that the dominant principal component was associated with TIRM⁻ and TIRM⁺ status, independently of age, sex and muscle type (multivariate co-efficient P-value TIRM⁻ = 0.0049, TIRM⁺ = 0.00722, Fig. 1D). The dominant muscle types (vastus lateralis and gastrocnemius medialis) were not associated with the dominant principal component. However, the single soleus and sartorius TIRM⁺ samples did separate from the remaining muscle samples by the dominant principal

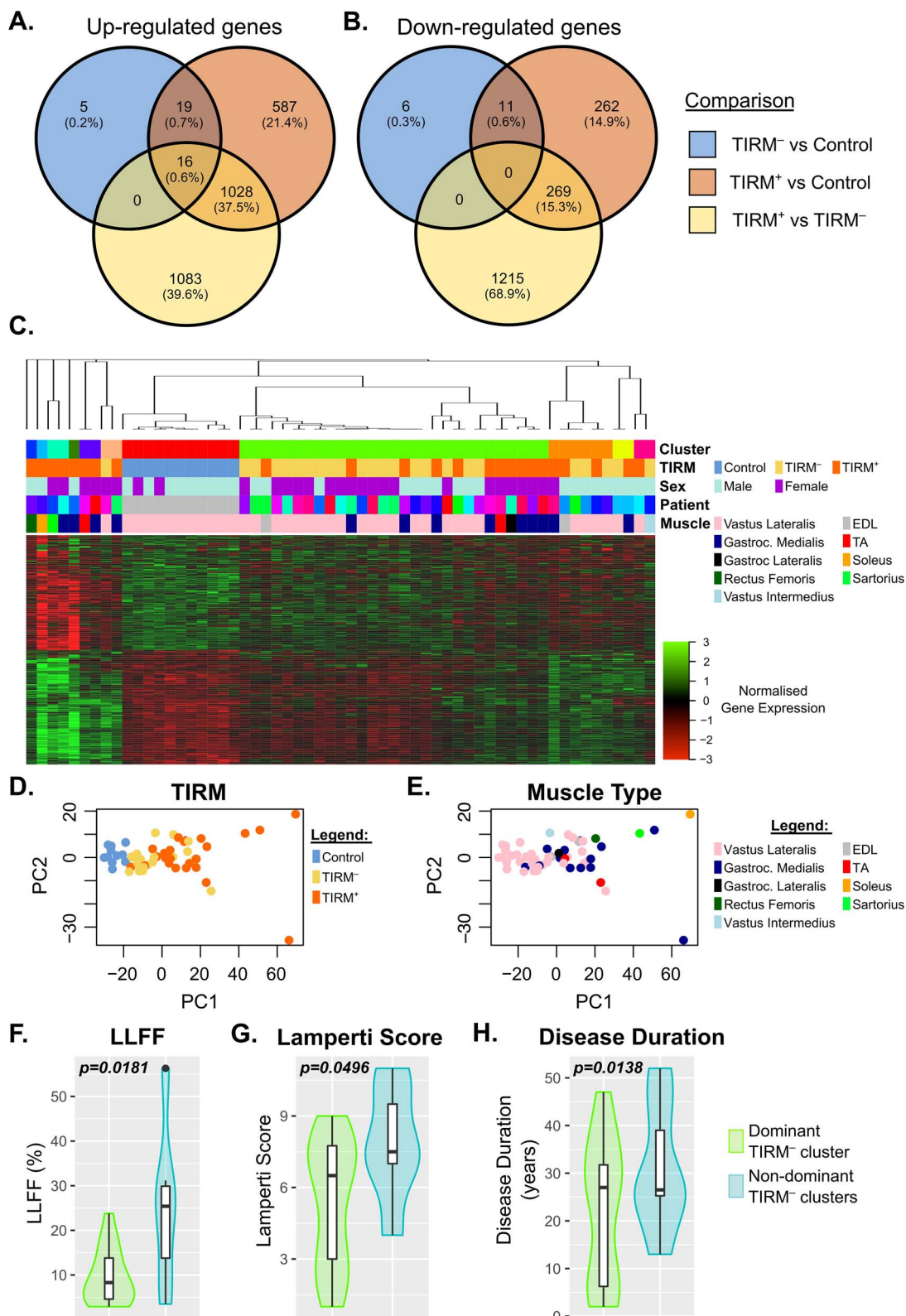


Figure 1. Differential gene expression analysis of muscle biopsies reveals clinically relevant FSHD subtypes. (A, B) Venn diagrams display overlaps in genes significantly (A) up-regulated and (B) down-regulated in comparisons between non-inflamed (TIRM⁻) FSHD muscle versus control (blue), inflamed (TIRM⁺) FSHD muscle versus control (orange) and TIRM⁺ vs TIRM⁻ FSHD muscle (yellow). (C) Heatmap displays the optimal 11 cluster solution of 56 muscle biopsy samples, employing 1015 genes significantly altered in TIRM⁻/TIRM⁺ FSHD muscle versus control as a feature space. z-normalized gene expression is coloured red to green and samples are labelled with TIRM and FSHD status, sex, muscle and patient identifier. (D, E) Scatter plots display samples according to the first and second principal components derived from the same 1015 gene feature space used for clustering. Samples are labelled with (D) TIRM status and (E) muscle type. (F–H) Boxplots display values of (F) LLFF, (G) Lamperti score and (H) disease duration for the 18 TIRM⁻ FSHD muscle biopsies which were assigned to the dominant FSHD cluster (green) compared to the 6 which lay outside this cluster (blue). Multivariate regression P-values assessing association with each clinical variable and dominant cluster status are displayed, adjusting for age, sex and muscle type.

component (multivariate co-efficient P -value soleus=0.0007, sartorius=0.025, Fig. 1E). There was no association between the dominant principal component and FSHD patient matched pair samples, indicating that inter-individual variability does not drive major variability in expression of genes in our feature space.

Importantly, one of the 11 clusters solely contained only control muscle biopsy samples, indicating that genes comprising the feature space are indeed perturbed in FSHD. Of the remaining 10 clusters containing only FSHD samples, the dominant cluster comprised 29 samples of which 62% (18 samples) are from the 24 TIRM⁻ samples and 38% (11 samples) are from the 24 TIRM⁺ samples. The remaining clusters contained either 6, 2 or 1 samples and are made up by the majority (54%, 13/24 samples) of TIRM⁺ samples and the 6 TIRM⁻ samples.

TIRM positivity indicates an inflammatory state associated with faster disease progression [50]. Given the imperfect separation of TIRM⁻ and TIRM⁺ samples, we postulated that TIRM⁺ samples in the dominant, majority TIRM⁻ cluster have milder clinical severity compared to FSHD samples outside this cluster. To address this, we first considered TIRM⁻ samples, 18 of which lay in the dominant cluster and 6 outside. Multivariate regression was performed to determine association between clinical severity and membership of the dominant cluster, using a variety of clinical assessments: Ricci clinical severity score, Lamperti clinical severity score, MRC sum score, LLFF, disease duration, MVC of TA and D4Z4 repeat length, adjusting for age, sex and muscle (where possible). The 18 TIRM⁻ samples in the dominant cluster displayed significantly lower LLFF (multivariate regression P -value = 0.0181, Fig. 1F) and Lamperti clinical severity scores (multivariate regression P -value = 0.0496, Fig. 1G) together with significantly shorter disease duration (multivariate regression P -value = 0.0138, Fig. 1H), compared to 6 TIRM⁻ samples outside this cluster, independently of patient age, sex and muscle. This is consistent with TIRM⁻ samples outside the dominant cluster being more clinically affected.

We next considered TIRM⁺ samples, 11 of which lay in the dominant cluster and 13 outside. There was no association between cluster membership and clinical variables for TIRM⁺ samples, after adjustment for age, sex and muscle.

Fibroadipogenic progenitor cell gene expression hallmarks FSHD muscle and discriminates TIRM⁺ from TIRM⁻ status

We next investigated the function of the genes differentially expressed in our FSHD samples via gene set enrichment analyses (GSEA). We found the FAP gene set defined by Rubenstein *et al.* [67] to be the most enriched gene set in the 40 genes up-regulated in TIRM⁻ FSHD muscle versus control (adjusted P -value = 1.3×10^{-5} Fig. 2A), and highly enriched in the 500 genes most significantly up-regulated when comparing TIRM⁺ muscle to control (adjusted P -value = 4.2×10^{-88} Fig. 2A) and TIRM⁺ muscle to isogenic TIRM⁻ muscle samples (adjusted P -value = 6.8×10^{-44} , Fig. 2A). This same gene set was also the most highly enriched among the 16 genes up-regulated in all three differential gene expression analyses and among these genes is the well characterized FAP marker PDGFRA (Fig. 2A). Additionally, several gene sets pertaining to FAP cell function in ECM deposition and maintenance were strongly enriched across comparisons (Fig. 2A).

Given this clear identification of gene expression signatures for FAPs in FSHD, we next assessed the relative contribution of FAPs to the gene expression landscape of each muscle biopsy using the deconvolution algorithm described by Rubenstein *et al.* [67] with the PLIER package in R [68]. As implied by the GSEA

results, we found that both TIRM⁻ and TIRM⁺ FSHD muscle biopsies had significantly higher FAP gene expression compared to control muscle biopsies independently of age, sex and muscle (multivariate P -value TIRM⁻ = 0.027, TIRM⁺ = 0.026, Fig. 2B). Moreover, isogenic TIRM⁺ biopsies had significantly higher FAP gene expression compared to TIRM⁻ biopsies from the same patient (P = 0.0018, Fig. 2B).

Elevated FAP gene expression in FSHD muscle has also been found in a previous study of largely non-isogenic muscle biopsies [55]. To investigate how consistent this finding is we analyzed a further 7 independent transcriptomic datasets profiling FSHD and control muscle biopsies, employing the same PLIER algorithm to determine FAP contribution in each sample. Strikingly, 6/7 datasets demonstrated significant up-regulation of FAP gene expression in FSHD samples compared to controls, and the seventh approached significance (P = 0.052) (Fig. 2C). This up-regulation was highly significant on meta-analysis (Fisher's combined P = 1.4×10^{-16} , Fig. 2C).

Genes associated with mitochondrial function and metabolism are suppressed in FSHD muscle

We next investigated genes suppressed in FSHD muscle via GSEA. There were no significantly enriched gene sets among the 17 genes suppressed in TIRM⁻ biopsies versus controls. GSEA on the 500 most down-regulated genes in TIRM⁺ biopsies versus controls demonstrated clear enrichment for mitochondrial genes and nucleic acid metabolism (Fig. 3A). GSEA on the 500 most down-regulated genes in the paired comparison of TIRM⁻ to TIRM⁺ biopsies also revealed strong enrichment for metabolic processes (Fig. 3B).

We have previously demonstrated insufficient mitochondrial biogenesis in differentiating human myoblasts from FSHD patients, which is linked to inadequate activation of ERR α by PGC1 α [69]. Given suppression of mitochondrial genes and metabolic processes in our new FSHD patient biopsies, we specifically examined ESRRA, encoding ERR α , and PPARGC1A, encoding PGC1 α . Expression of both ESRRA and PPARGC1A was suppressed in TIRM⁺ muscle compared to control (Fig. 3C and D), confirming our *in vitro* findings [69] in FSHD patients.

Additionally, total expression of all 13 protein-coding genes in the mitochondrial genome, a metric shown to correlate with mitochondrial DNA copy number in skeletal muscle [70] and several types of cancer [71], was significantly suppressed in both TIRM⁻ and TIRM⁺ samples compared to controls independently of age and sex. This indicates lower mitochondrial content in FSHD samples (Fig. 3E).

Differential transcript analysis of FSHD muscle biopsies demonstrates perturbation of autophagy and PAX7 target genes

To examine alternative splicing, we next performed differential transcript analyses using DRIMSeq [72] and *stageR* [73] packages in R. Adjusting for matched pairs imposed strict restrictions on the transcript distribution across samples, limiting the number of genes we could assess in a manner which may introduce bias. We therefore focused on comparisons of control versus TIRM⁻ or TIRM⁺ status, adjusting for age and sex of participants. We followed an established protocol to identify genes which displayed evidence of both differential transcript usage (DTU) and differential transcript expression (DTE) [74].

We identified 64 genes with evidence of DTU and DTE in TIRM⁻ samples versus controls and 121 genes in TIRM⁺ samples versus

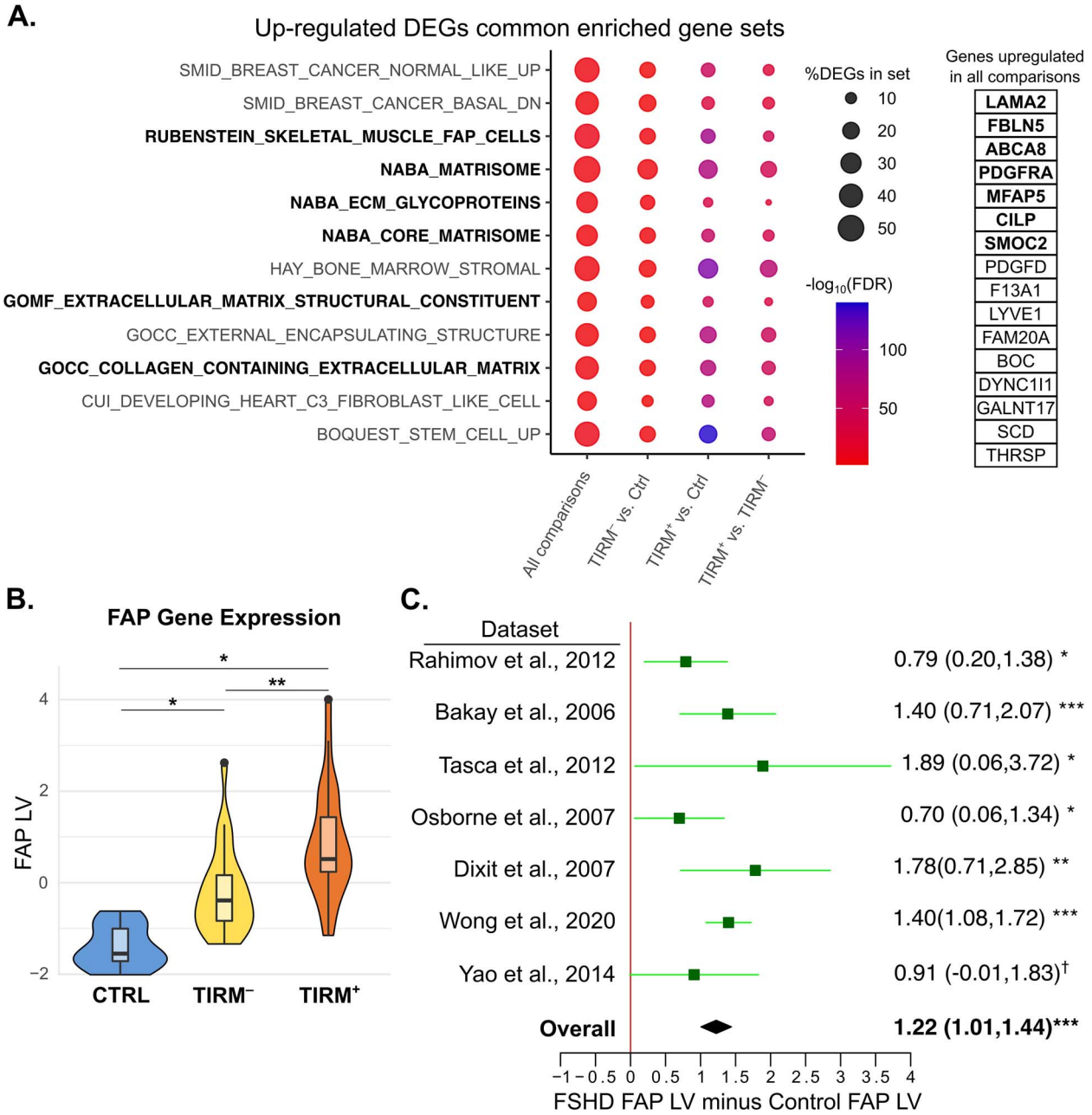


Figure 2. Fibroadipogenic progenitor (FAP) gene expression is up-regulated in FSHD muscle biopsies and increases with inflammation. (A) Dot plot heatmap displays gene sets in MSigDB which were consistently enriched among differentially expressed genes from four comparisons: The 16 genes up-regulated in all comparisons (listed beside the plot), all genes up-regulated in TIRM⁻ versus control, the 500 most significant genes up-regulated in TIRM⁺ versus control, the 500 most significant genes up-regulated in TIRM⁺ versus TIRM⁻ FSHD muscle. The size of the dot is proportional to the percentage of up-regulated genes found in each enriched gene set, dots are coloured red to blue according to $-\log_{10}(\text{FDR})$. The FAP gene set is highlighted in bold text, as are the 7/16 genes up-regulated in all comparisons which overlap with the FAP gene set. Gene sets related to ECM maintenance are also in bold text. (B) Boxplot displays the value of the latent variable associated with FAPs and no other muscle biopsy cell type (FAP LV) following deconvolution of our 56 muscle biopsy samples employing the approach of Rubenstein et al. [67]. Samples are separated into control (blue), TIRM⁻ (yellow) and TIRM⁺ (orange). An asterisk denotes a significance of $P < 0.05$ and two asterisks denotes $P < 0.01$ from multivariate regression determining association of the FAP LV with control vs TIRM⁻ or TIRM⁺ status independently of age and sex, or a paired comparison of matched TIRM⁻ and TIRM⁺ samples. (C) Forrest plot displays the difference between FAP LV values in FSHD and control samples, following deconvolution of 7 independent FSHD muscle biopsy microarray or RNA-seq datasets (130 FSHD, 98 control). In 6/7 datasets and on meta-analysis, the FAP LV values are significantly elevated on FSHD samples, implying increased FAP gene expression in FSHD. Boxes denote the mean difference in FAP LV between FSHD and control muscle biopsies and whiskers denote 95% confidence interval. A vertical line denotes a FAP LV difference of 0 and datasets where the whiskers cross this line have not attained significance at $P < 0.05$ (as assessed by Wilcoxon test). Values for mean FAP LV difference and confidence interval are displayed for each dataset with significance level, where an asterisks denotes $P < 0.05$, two asterisks denotes $P < 0.01$, three asterisks denotes $P < 0.001$, and † denotes $P = 0.052$. The overall estimate is displayed as a diamond and was computed using a random effects model with significance assessed via Fisher’s combined test.

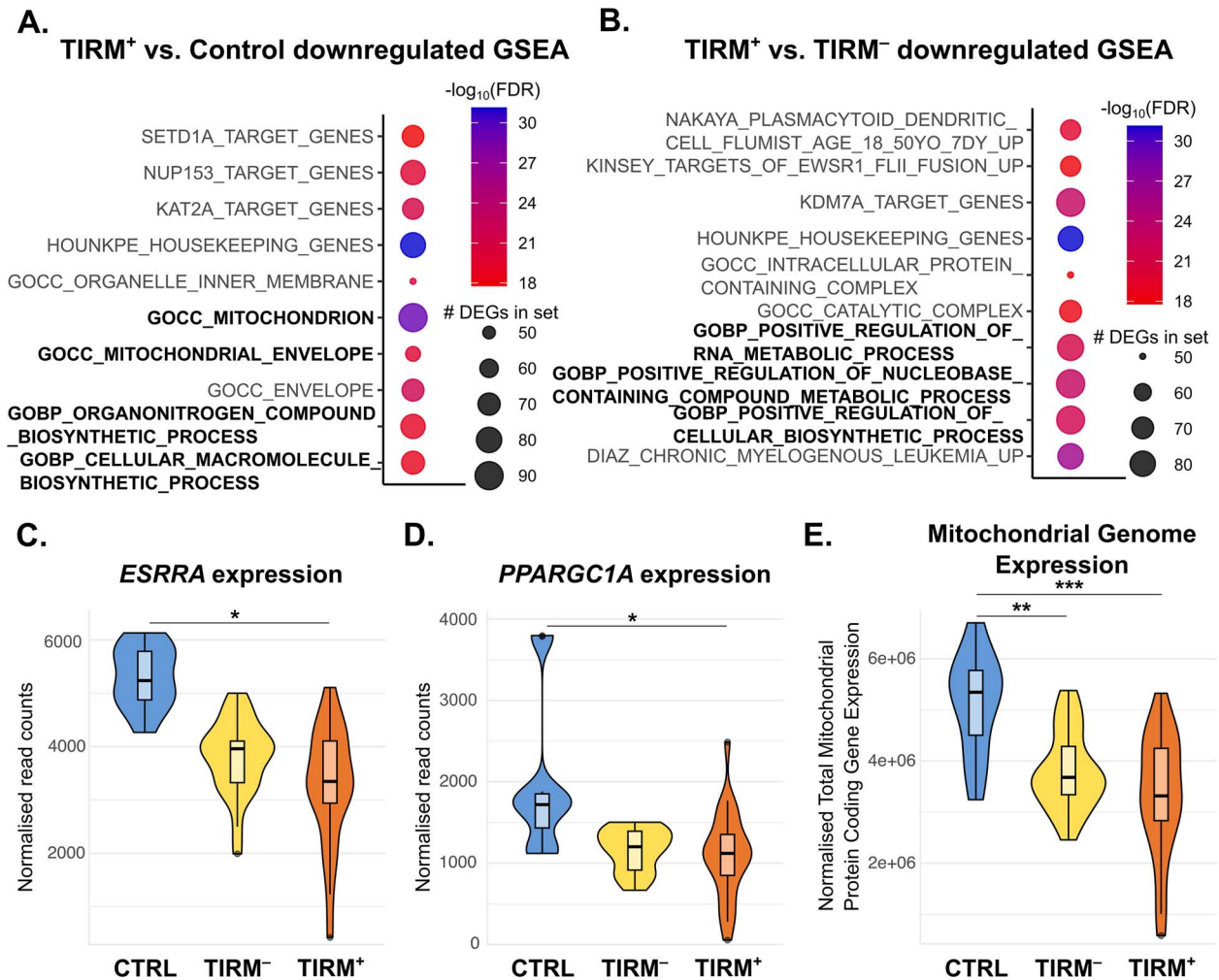


Figure 3. Mitochondrial gene expression and key biogenesis regulators are suppressed in FSHD muscle. (A, B) Dot plot heatmaps displays gene sets in MSigDB which were most significantly enriched among (A) the 500 most significant genes down-regulated in TIRM⁺ versus control and (B) the 500 most significant genes down-regulated in TIRM⁺ versus TIRM⁻ FSHD muscle. The size of the dot is proportional to the number of down-regulated genes in each enriched gene set, dots are coloured red to blue according to $-\log_{10}(\text{FDR})$. Mitochondrial and metabolic gene sets are in bold text. (C, D) Boxplots display normalized read counts for (C) *ESRRR* and (D) *PPARGC1A* across control (blue), TIRM⁻ (yellow) and TIRM⁺ (orange) samples, where an asterisk denotes adjusted $P < 0.05$ from DESeq analysis of differential expression, adjusting for age, sex and matched pair. (E) Boxplot of the sum of normalized read counts for the 13 protein coding mitochondrial genes across control, TIRM⁻ and TIRM⁺ samples. Two asterisks denotes $P < 0.01$, and three asterisks $P < 0.001$ from multivariate regression determining the association of mitochondrial genome expression with control vs TIRM⁻ or TIRM⁺ status independently of age and sex.

controls, with an overlap of 25 genes in both comparisons (Supplementary Table S2). GSEA on the 121 genes alternatively spliced in TIRM⁺ muscle versus control revealed strong enrichment for autophagic processes, immune processes and PAX7 target genes, an interesting finding considering PAX7 target gene are repressed in FSHD muscle [26, 35, 36] (Fig. 4A). The 64 genes alternatively spliced in TIRM⁻ muscle versus control were enriched for a number of processes including target genes of both BMP2, a signalling protein that inhibits adipogenic differentiation of FAP cells [75], and the nuclear DNA-binding protein HMGB1, which modulates muscle inflammation [76] and regulates PAX7 expression [77] during myogenesis (Fig. 4B).

Among genes alternatively spliced in TIRM⁺ muscle versus control is *TNNT1*, the slow muscle troponin, which has two main isoforms showing signs of differential usage: ENST00000356783 (*TNNT1-202*), which lacks 11 amino acids encoded by exon 5 of the higher molecular weight ENST00000291901 (*TNNT1-201*). We see specific down-regulation of the high molecular weight *TNNT1-201* in TIRM⁺ FSHD muscle compared to control,

with a trend to up-regulation of the low molecular weight *TNNT1-202* (Fig. 4C). We designed primers to distinguish these two isoforms and confirmed down-regulation of the high molecular weight *TNNT1-201* in TIRM⁺ FSHD muscle compared to control by RT-qPCR in a subset of our muscle biopsy samples (Fig. 4D).

A second example of an alternatively spliced gene is *SAMD4A*, encoding the RNA-binding translational repressor SMAG1 protein, which exhibits compensatory DTE in two transcript isoforms between control and TIRM⁺ samples. TIRM⁺ samples up-regulate ENST00000392067, which encodes the canonical protein, while control samples up-regulate ENST00000555719 (*SAMD4A-209*), a severely truncated isoform consisting of 3 exons (Fig. 4E). Again, we designed primers to distinguish these two isoforms and confirmed up-regulation of the transcript encoding canonical SMAG1 in TIRM⁺ FSHD muscle compared to control by RT-qPCR in a subset of our muscle samples (Fig. 4F).

Among genes exhibiting differential splicing patterns in TIRM⁻ muscle versus control was *MYH14*, a non-muscle myosin heavy

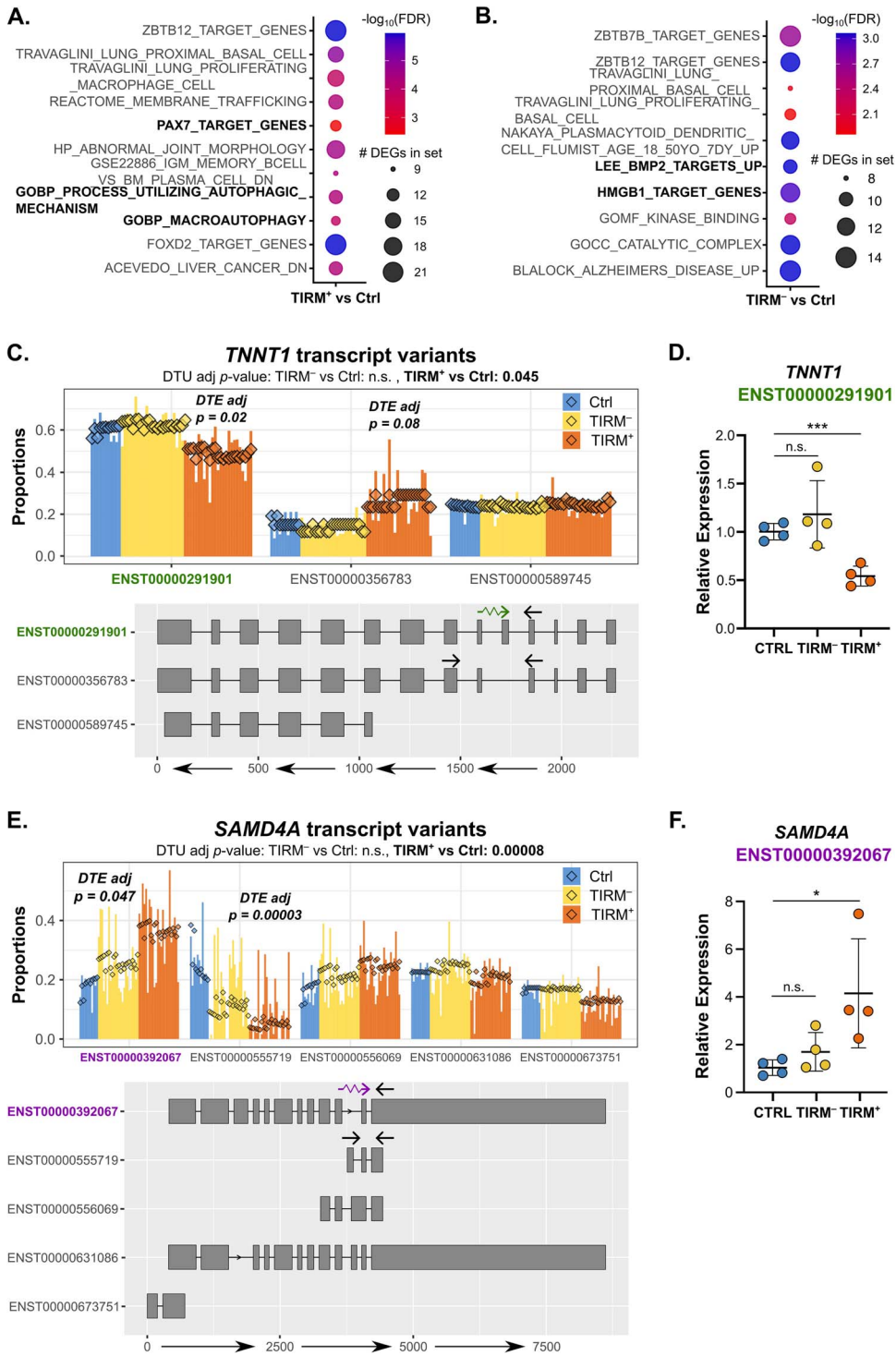


Figure 4. Alternative splicing analysis in muscle biopsies reveals novel FSHD changes. (A, B) Dot plot heatmaps displays gene sets in MSigDB which were most significantly enriched among genes alternatively spliced in (A) TIRM⁺ versus control and (B) TIRM⁻ versus control. The size of the dot is proportional to the number of alternatively spliced genes found in each enriched gene set, dots are coloured red to blue according to $-\log_{10}(\text{FDR})$. PAX7 target genes, autophagic processes and BMP2 and HMGB1 target genes in bold text. (C–F) Bar plots generated using our bespoke software (Supplementary File S1) display the proportion of total gene expression for (C) TNNT1 and (E) SAMD4A accounted for by the major transcript variants, across 11 control samples (blue), 24 TIRM⁻ FSHD samples (yellow) and 24 TIRM⁺ FSHD samples (orange). Transcript structures are displayed beneath each plot and coloured text highlights and explains the transcript that is significantly altered in its expression in TIRM⁻ (yellow) or TIRM⁺ (orange). At the top of each plot adjusted *P*-values are displayed denoting the significance of differential transcript usage (DTU) for the gene between control and TIRM⁻ and control and TIRM⁺ samples (adjusted for age, sex and multiple comparisons). On each plot is displayed the adjusted *P*-value for differential transcript expression (DTE) of the highlighted transcript in the comparison for which DTU is significant, adjusting for age, sex and multiple comparisons. Validation of (D) TNNT1 and (F) SAMD4A expression of highlighted isoforms was performed by quantitative real-time PCR in 4 control samples (blue), 4 TIRM⁻ samples (yellow) and their paired isogenic TIRM⁺ samples (orange), using RNA from samples selected on predicted DTE and quantity of RNA available. TNNT1 ENST00000291901 expression was quantified relative to sum expression of ENST00000291901 and ENST00000356783, and SAMD4A ENST00000555719 expression was quantified relative to that of ENST00000392067, the other isoform with DTE. Primer locations are indicated by arrows on the transcript structures in (C) and (D). Data is presented as $\Delta\Delta\text{Ct}$ relative to controls, with each point representing the mean of 3 technical replicates per biopsy. Black bars represent group mean \pm SD, and significance is indicated by an asterisk that denotes $P < 0.05$ or 3 asterisks that denote $P < 0.01$.

chain, with 2 main spliced isoforms: ENST00000376970 (NMHCII-C0) which lacks 8 amino acids within the globular head domain encoded by exon 6, compared to ENST0000042560 (NMHCII-C1). We found up-regulation of NMHCII-C0 specifically in FSHD TIRM⁻ muscle compared to control (Supplementary Fig. S1A). A second gene exhibiting DTU in TIRM⁻ muscle compared to controls is RFX5, a transcriptional activator of major histocompatibility complex (MHC) class II promoters. Compared to controls, TIRM⁻ samples exhibit up-regulation of ENST00000412774 (RFX5-204), a severely truncated isoform lacking sequences encoding most of the N-terminus and part of the C-terminus (Supplementary Fig. S1B).

To facilitate independent investigation of the splicing patterns of other genes in TIRM⁻ and TIRM⁺ FSHD muscle compared to control, we have developed a data visualization tool, written in shiny R [78]. The programme displays the transcript structure, relative expression in control, TIRM⁻ and TIRM⁺ muscle biopsies and corresponding P-values for any given gene with significant DTU in our above comparisons (Supplementary File S1).

Differential gene expression of FSHD and control peripheral blood mononuclear cells (PBMCs) demonstrates suppression of antigen processing and presentation

RNA-sequencing had also been performed on peripheral blood mononuclear cells (PBMCs) that were isolated from 14 FSHD1 patients, 1 FSHD2 patient and 14 control patients [60]. We have previously demonstrated no difference in age or sex distribution between control and FSHD PBMC samples in this cohort [60]. Differential expression analysis was performed analysing length adjusted transcript counts aggregated to gene level, via negative binomial models using the DESeq2 package [65], to determine gene expression associated with FSHD PBMCs, adjusting for age and sex.

Gene expression differences between FSHD and control PBMCs were more subtle than muscle and we identified only 26 genes up-regulated and 21 genes down-regulated in FSHD PBMCs compared to controls after adjusting for multiple comparisons. Of these 47 genes, only 16 were expressed in more than 3 samples. Using expression of these 16 genes as a feature space, K-medoids clustering of the 29 PBMC samples revealed a 7 cluster solution with poor separation of FSHD and control samples (Fig. 5A), despite the dominant principal component associating with FSHD status, independently of age and sex (multivariate regression $P = 9.7 \times 10^{-9}$, Fig. 5B).

GSEA on the 26 genes up-regulated in FSHD PBMCs demonstrated no enriched gene sets. However, GSEA on the 21 genes down-regulated in FSHD PBMCs demonstrated enrichment for antigen processing and presentation as well as adaptive immunity (Fig. 5C).

As per muscle, DUX4 was again not differentially expressed between FSHD and control samples. However, in contrast to muscle biopsies, DUX4 transcripts were expressed in 13/15 FSHD patient and 11/14 control PBMC samples. Expression of DUX4 in PBMCs was low but higher than seen in muscle biopsies (max 8 reads per sample in FSHD and 6 in controls, Fig. 5D).

Differential transcript expression but not differential gene expression permits near perfect separation of FSHD and control PBMCs

We next performed DTU and DTE analysis to identify alternative splicing events in FSHD PBMCs compared to controls, adjusting for age and sex of participants. We identified 36 genes with evidence

of DTU and DTE in FSHD PBMCs. GSEA on these genes revealed no significantly enriched gene sets.

The 36 genes yielded 40 transcripts with evidence of differential expression between FSHD and control samples. Given that differentially expressed genes facilitated poor separation of FSHD and control PBMCs via clustering analysis, we repeated clustering analysis using expression of these 40 transcripts as a feature space. K-medoids clustering yielded a 2 cluster solution which almost perfectly separated FSHD and control samples, with the exception of a single control sample (Fig. 5E). This indicates that the FSHD state is better discriminated in PBMCs by consideration of differential transcript profiles than differential gene expression.

As with the muscle biopsies, to facilitate independent investigation of splicing patterns in PBMCs we have developed a data visualization in the shiny package [78] in R (Supplementary File S2).

Discussion

We have investigated differential gene expression, differential transcript usage and differential transcript expression in isogenic non-inflamed (TIRM⁻) and inflamed (TIRM⁺) FSHD muscle biopsies alongside matched control muscle biopsies, as well as in FSHD and control PBMCs.

FSHD is highly heterogeneous, with monozygotic twins often progressing at dramatically different rates [79]. This unpredictable heterogeneity is a key stumbling block to FSHD clinical trials, where for example, unintentional overloading of treatment arms with fast progressors could mask detection of treatment efficacy. In similarly heterogeneous pathologies such as breast cancer, molecular subtyping is extremely useful in identifying homogenous subgroups of patients, facilitating personalized medicine approaches [80, 81]. D4Z4 repeat length has been linked with accelerated progression in FSHD [82] as has male sex [83], and MRI studies show that inflammation can accelerate clinically relevant fatty replacement in FSHD muscle [50]. We have also demonstrated that FSHD patients can follow a number of distinct clinical phenotypes, independently of sex and D4Z4 repeat length [84]. However, clinically relevant molecular subtypes that provide the basis of a testable, quantitative grouping have not been described in FSHD.

Here we show that expression of 1015 genes, differentially expressed in TIRM⁻ and TIRM⁺ FSHD muscle compared to controls, separates our samples into a control cluster and 10 distinct FSHD clusters. The largest FSHD cluster contains TIRM⁻ muscle samples which are clinically milder than TIRM⁻ muscle samples in the other FSHD clusters, with lower Lamperti clinical severity scores, lower LLFF and shorter disease duration, despite equal D4Z4 repeat length and independent of age and sex. This is the first evidence that clinically relevant molecular subtypes of FSHD exist, motivating investigation of molecular subtype biomarkers in larger cohorts.

FAPs are interstitial muscle-resident multipotent progenitor cells, with capacity to differentiate into myofibroblasts, adipocytes and, under certain conditions, chondrocytes and osteocytes [75]. FAPs are activated and proliferate after muscle injury and differentiate following specific molecular cues into specialized cells that prepare the extra-cellular matrix to support muscle regeneration. Mis-regulation of FAP function has been implicated in aberrant muscle regeneration in several neuromuscular disorders including LGMD2B [85], ALS [86] and recently FSHD [53, 55, 58]. FAP dysregulation in FSHD was first

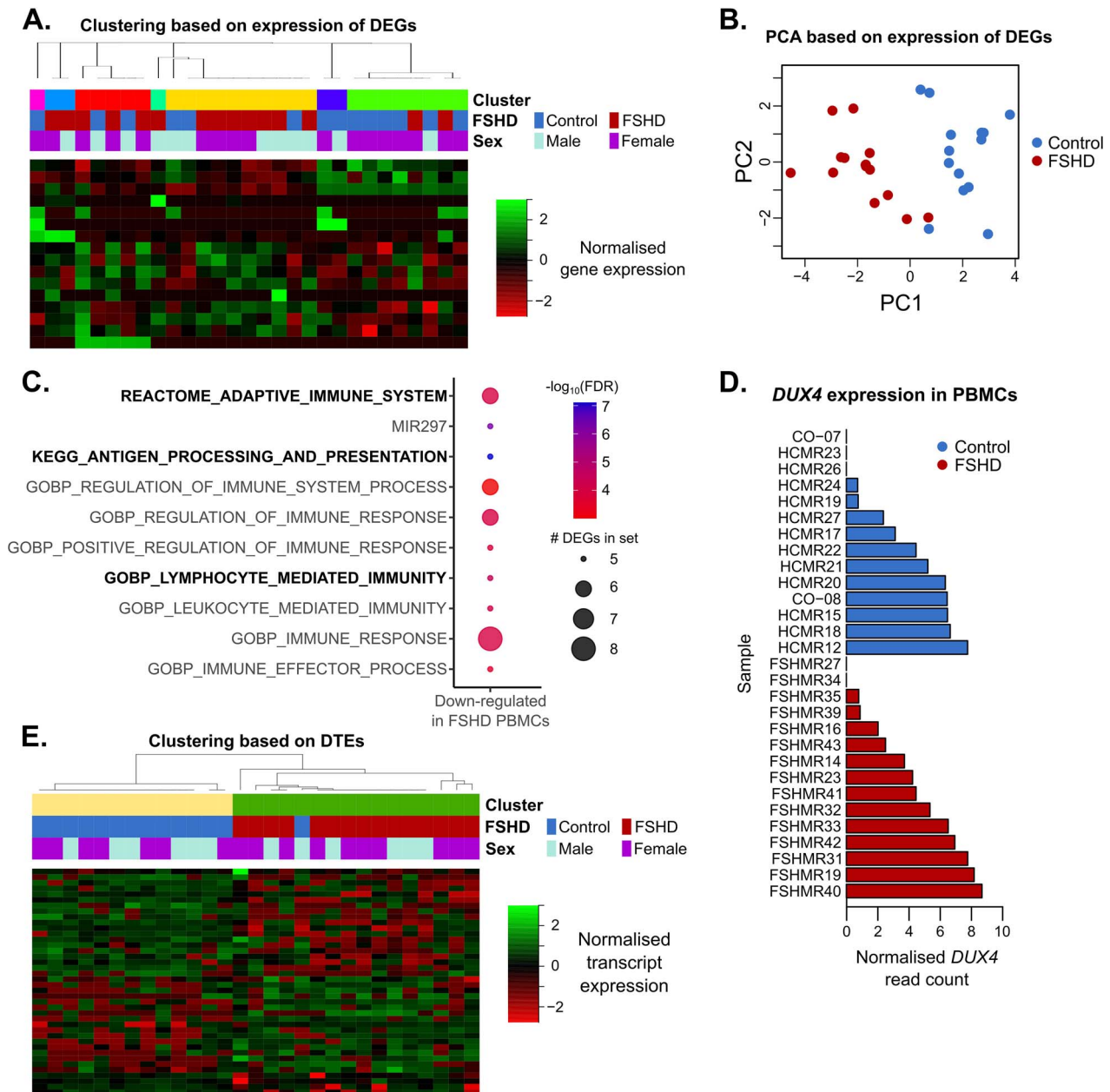


Figure 5. FSHD PBMCs display suppression of antigen presentation and consistent alternative splicing features. (A) Heatmap displays the optimal 7 cluster solution of 29 PBMC samples, employing 16 genes significantly altered in FSHD vs control PBMCs, expressed in >3 samples as a feature space. z-normalized gene expression is coloured red to green and samples are labelled with FSHD status and sex. (B) Scatter plot display samples according to the first and second principal components derived from the same 16 gene feature space used for clustering, labelled with FSHD status. (C) Dot plot heatmap displays gene sets in MSigDB which were most significantly enriched among genes suppressed in FSHD PBMCs. The size of the dot is proportional to the number of suppressed genes found in each enriched gene set, dots are coloured red to blue according to $-\log_{10}(\text{FDR})$. Adaptive immunity gene sets are highlighted in bold text. (D) Bar plot displays normalized read count for *DUX4* in each control (blue) and FSHD (red) PBMC sample. (E) Heatmap displays the optimal 2 cluster solution of 29 PBMC samples, employing 40 transcripts with evidence for DTE within genes with evidence for DTU in FSHD vs control PBMCs as a feature space. z-normalized transcript expression is coloured red to green and samples are labelled with FSHD status and sex.

noted in a mouse model of stochastic transient *DUX4* expression, which resulted in an increased FAP gene expression signature [58] that was validated histologically and found to associate with muscle fibrosis post injury [87]. Subsequently, an increased FAP transcriptomic signature was found in a dataset of 39 mainly non-isogenic FSHD muscle biopsies, where it correlated with FSHD transcriptomic biomarkers of severity [55]. Recently a detailed characterization of FSHD patient mesenchymal cells demonstrated increased numbers of CD201⁺ and PDGFRA⁺ FAPs in FSHD, which correlated with accelerated fibrosis and disease

progression and showed inhibited adipogenesis *ex vivo* [53]. We found that FAP gene expression is the most enriched up-regulated gene set in FSHD muscle versus control. We analyzed isogenic samples and demonstrated that within patients, inflamed TIRM⁺ muscle displayed elevated levels of FAP gene expression to paired TIRM⁻ biopsies. Importantly, the FAP marker *PDGFRA* is up-regulated in TIRM⁻ and TIRM⁺ FSHD muscle versus control, as well as between matched TIRM⁺ and TIRM⁻ FSHD muscle. We confirmed this finding by performing meta-analysis of multiple independent FSHD published transcriptomic datasets,

demonstrating that elevated FAP gene expression is a clear feature of FSHD muscle across 130 FSHD samples.

Mitochondrial dysfunction is also a well-studied feature of FSHD muscle [21, 33, 34, 69]. We have previously demonstrated that FSHD patient-derived myoblasts exhibit a regeneration defect characterized by small, hypotrophic myotubes [34, 69]. This is driven by suppression of mitochondrial biogenesis consequential of *ERR α* /*PGC1 α* down-regulation [69], and dysfunctional mitochondria with elevated mitochondrial membrane potential and increased mitochondrial ROS generation [34]. *ERR α* agonist food supplements such as biochanin A, genistein and daidzein [69], as well as mitochondrially targeted anti-oxidants [34], can rescue such defective in vitro FSHD myogenesis. Here we show that mitochondrial genes are significantly suppressed in inflamed TIRM⁺ FSHD muscle, providing a direct patient level correlate of our in vitro finding. In particular genes encoding *ERR α* and *PGC1 α* are significantly down-regulated in TIRM⁺ muscle, motivating translation of *ERR α* targeted food supplements to the clinic [69]. Reduced mitochondrial genome expression in both TIRM⁻ and TIRM⁺ FSHD samples further implies a progressive suppression of mitochondrial content in FSHD muscle.

Increased FAP-related gene expression and reduced mitochondrial gene expression in TIRM⁻ FSHD biopsies compared to muscle-matched controls indicates a baseline dysregulation of these processes in FSHD muscle regardless of inflammation.

The TIRM⁻ and control muscle biopsies were from the quadriceps femoris (with 34/35 being from vastus lateralis). However, since inflammation in FSHD muscle is sporadic and highly heterogeneous between patients, it is rare to identify patients from which contralateral TIRM⁻ and TIRM⁺ samples can be taken from the same muscle. As such, the isogenic TIRM⁺ muscle samples come from a range of muscle types (although 15/24 were from gastrocnemius), raising the concern that gene expression changes in TIRM⁺ samples are due to transcriptional differences between different muscles [88]. While we cannot completely adjust for this covariate to eliminate the possibility of a muscle type contribution, principal component analysis on the genes used to identify expression patterns linked to TIRM status (such as FAPs and mitochondrial biogenesis) demonstrated strong association of the dominant principal component with TIRM status independently of muscle type (Fig. 1D and E). This indicates that differences in gene expression are more likely associated with TIRM status than muscle type in FSHD. Furthermore, several studies investigating human lower limb muscle transcriptomes have shown that variability in gene expression between individuals is substantially greater than that between muscle types within an individual [88, 89]. Thus for phenotypes such as inflammation, the noise contribution will be lower when comparing different muscle types within an individual than comparing the same muscle type across multiple individuals. Interestingly, such studies have also reported that certain muscles in healthy individuals, including those more commonly affected in FSHD (e.g. gastrocnemius), exhibit increased FAP gene expression and mitochondrial down-regulation compared to muscles such as the vastus lateralis [88]. This suggests that certain muscle groups may be predisposed to being more susceptible to pathological changes driving FSHD, which may explain the highly specific, but poorly understood, FSHD-affected muscle distribution.

TIRM⁻ FSHD muscle also down-regulated genes associated with RNA metabolism and nucleobase metabolism, a feature previously observed in *DUX4* over-expressing myotubes, where it correlated with inhibited nonsense-mediated decay [90] and aberrant RNA splicing [41]. Thus, we performed the first

genome wide analysis of differential transcript usage and expression in FSHD muscle. Genes with evidence of differential transcript expression in inflamed FSHD muscle were enriched for autophagic processes, of significance given an elevated level of autophagic vacuoles found in a case series of atypical FSHD patients [91], and a potential role for *DUX4* in autophagy [92]. Curiously, enrichment for alternative splicing of *PAX7* target genes in inflamed FSHD muscle biopsies was also observed. We have demonstrated, and others have confirmed, that *PAX7* target genes are significantly suppressed in FSHD muscle [26, 55, 56], correlating with disease severity and progression [35, 36, 60]. It is possible that alternative splicing of *PAX7* target genes drives exon suppression which in turn leads to *PAX7* target suppression at the gene level, however dedicated studies would be necessary to test this hypothesis. Genes with differential transcript expression in TIRM⁻ FSHD muscle compared to control demonstrated enrichment of *BMP2* and *HMGB1* signalling. *BMP2* drives osteogenic differentiation of FAPs while inhibiting adipogenic differentiation [75]. FAPs isolated from FSHD patients show inhibited adipogenic differentiation [53], while *DUX4* expression is associated with osteogenesis [93]. *HMGB1* is a critical mediator of muscle regeneration and inflammation in multiple muscular dystrophies [76]. Importantly, fully reduced *HMGB1* contributes to muscle regeneration, while oxidized *HMGB1* (e.g. by ROS) is a pro-inflammatory cytokine. *HMGB1* also shares high structural homology with closely related family member *HMGB2* [94], which is part of an inhibitory DNA binding complex that binds to the D4Z4 region at the *DUX4* promoter [95]. Given that FSHD is a pathology characterized by elevated intracellular ROS [34], inhibited regeneration [54] and inflammation [50], further investigation of *HMGB1* in FSHD is ongoing.

Several post-transcriptional misregulation events that we observed in FSHD muscle have precedent in pathologically relevant contexts. The precise splicing pattern of *MYH14* observed in TIRM⁻ samples compared to controls is also observed in Myotonic Dystrophy type 1 (DM1), a repeat expansion myopathy with associated sensorineural deafness [96]. *MYH14* is mutated in an autosomal dominant sensorineural deafness [97] and plays a role in mitochondrial fission [98], and both sensorineural deafness [6] and mitochondrial dysfunction [34, 69] are features of FSHD. The alternative splicing pattern of *TNNT1* observed in TIRM⁺ samples compared to controls resembles a pattern reported in muscle from patients with Charcot-Marie-Tooth disease type 1 (CMT1), but not CMT type 2, where it associates with increased specific force in individually isolated muscle fibres, as well as chronically overused muscle from a prior Polio patient [99]. The same splicing pattern of *TNNT1* has also been shown in muscles of trained older adults compared to sedentary participants, and again associated with increased fibre force independent of cross-sectional area [100].

Loss of *SAMD4A* has been linked to changes in RNA-binding and post-transcriptional regulation that can exacerbate myopathy [101] and cause metabolic defects such as uncoupling of mitochondrial respiration [102]. Increase in canonical *SAMD4A* in TIRM⁺ samples, rather than the truncated and likely non-functional isoform up-regulated in control samples, may reflect counteractive measures against metabolic defects affecting FSHD muscle. Finally, *RFX5* is an essential member of the RFX complex, which binds to the X-box sequence in the promoter of MHC-II genes and facilitates transcriptional activation of MHC-II genes through association with other regulatory factors such as NF-Y [103, 104]. The *RFX5* isoform over-expressed in TIRM⁻ samples lacks part of the DNA-binding domain as well as a C-terminal

sequence necessary for cooperative binding, both of which are necessary for activation of MHC-II gene expression [104]. This is interesting given the enrichment of adaptive immunity pathways in genes differentially expressed between control and FSHD PBMCs, particularly considering a report connecting DUX4 expression in tumours to suppression of both MHC-I and MHC-II related gene expression [105, 106].

Together, these splicing patterns indicate that FSHD muscle may be adapting to a chronic pathogenic environment to improve muscle contractility, and that these adaptations are occurring in inflammatory stages of disease pathology prior to onset of fat infiltration. Deeper investigation of the remaining 183 alternatively spliced genes (Supplementary File S1) is ongoing to identify additional alternative splicing events with clinical relevance. While markers identified here in terms of gene expression and alternative splicing are limited to earlier, inflamed stages of muscle pathology, similar analyses in patient datasets containing a subset of muscles with fat infiltration [52, 55, 56] will inform whether these patterns observed persist through disease progression and may identify further transcriptional and/or post-translational perturbations affecting later stages of pathology.

Analysis of FSHD and control PBMCs demonstrated few differentially expressed genes, however genes suppressed in FSHD PBMCs were enriched for antigen presentation and processing. This is interesting given that DUX4 can suppress MHC class I expression in several malignancies in a manner associated with tumour immunotherapy resistance [106]. Suppression of antigen processing in FSHD PBMCs may therefore be reflective of a suppressed MHC class I presentation on DUX4 expressing cells throughout the patient such as skeletal muscle, making this an area of interest for future work. Curiously, DUX4 expression was higher in PBMCs than muscle biopsies, and both control and FSHD PBMCs expressed DUX4 transcripts. This is in line with the detection of DUX4 in immortalized lymphoblastoid cell lines from both FSHD and control individuals [48, 107]. Lastly, FSHD PBMCs showed evidence of differential transcript usage and expression compared to control PBMCs, and clustering based on differential transcript level, led to a more accurate separation of FSHD and control PBMCs compared to differential gene expression.

For ready investigation of the splicing patterns of other genes in TIRM⁻ and TIRM⁺ FSHD muscle and FSHD PBMCs compared to control, we provide data visualization tools (Supplementary Files S1 and S2). The programmes display the transcript structure, relative expression in control, TIRM⁻ and TIRM⁺ muscle biopsies and FSHD PBMCs and corresponding P-values for any given gene with significant DTU.

In summary, we performed a detailed transcriptomic analysis of isogenic FSHD non-inflamed and inflamed skeletal muscle as well as FSHD PBMCs. Our analyses indicate FAP cell population expansion and mitochondrial dysfunction in FSHD muscle, which progress in the presence of inflammation. Our transcript-level investigation provides global analysis of RNA splicing in FSHD muscle, detecting differences in post-transcriptional regulation with potential pathological relevance including BMP2 and HMBG1, which were previously masked by gene-level aggregation. Finally, we describe defects in antigen presentation in FSHD PBMCs indicating a potential role for DUX4 in systemic MHC I expression dynamics and propose transcript-level analysis as a powerful approach to blood based biomarker discovery.

Materials and Methods

Patient cohort

We have described the patient cohort before [60] but provide details here for reference. We considered data corresponding to 49 individuals: 26 FSHD patients (n = 25 FSHD1 and 1 FSHD2) and 23 unrelated unaffected control individuals, all investigated between 2019–2021 at the Radboud University Medical Center (Nijmegen, The Netherlands) Neurology Outpatient clinic. Subjects provided written informed consent and the study received regional medical ethical committee approval (CMO Arnhem-Nijmegen). Clinical data collected included: age of onset, FSHD type, D4Z4 repeat length, age of onset, disease duration, and LLFF assessed by MRI. Patients underwent muscle power assessment via the MRC sum score [108], in which muscle power grade (rated 0–5 per muscle) was assessed for 6 muscle groups on left and right, giving a sum score out of 60. FSHD patients also underwent the three additional clinical assessments: 1) the Ricci score [63], which ranges from 0–10 (0 = no symptoms, 10 = wheelchair bound) and assumes a typical progression of muscle weakness descending from face to legs, 2) the Lamperti score [64], a 0–15 severity scale (0 = low, 15 = high) which independently evaluates weakness of 6 muscle regions affected in FSHD to better account for atypical cases, and 3) MVC of tibialis anterior [109] when able. We have previously analyzed clinical variables and age and sex distributions in this dataset, showing Ricci, Lamperti and MRC sum scores are elevated in older, male FSHD patients [60]. FSHD patients who underwent muscle biopsy were significantly older than controls and showed a male sex bias [60]. A full overview of clinical data collected is provided in our previous publication [60].

MRI acquisition

MRI imaging used to select muscle biopsy site was provided in detail previously [60] using a protocol established by Mul et al. [110]. To reiterate, Dixon sequences were analyzed using MATLAB and ImageJ software. MATLAB was used to calculate a fat fraction (FF) map from the water and fat image of the Dixon sequence according to: $FF = \frac{F}{F+W}$. Contours of twelve upper leg muscles (sartorius, gracilis, vastus medialis, vastus lateralis, vastus intermedius, rectus femoris, biceps femoris caput brevis, biceps femoris caput longus, semitendinosus, semimembranosus, adductor magnus, adductor longus) and seven lower leg muscles (tibialis anterior, extensor digitorum longus, peroneus, tibialis posterior, soleus, gastrocnemius medialis, and gastrocnemius lateralis) were outlined manually on the FF map as described [110], and average FF within a single muscle contour (\overline{FF}_i) was calculated in ImageJ. Average Lower limb fat fraction (LLFF) was calculated as follows: $LLFF = \frac{\sum_{i=1}^n A_i \overline{FF}_i}{\sum_{i=1}^n A_i}$, where A = area of each muscle and n = muscle contours of one patient.

Biopsy collection

As previously detailed [60], muscle biopsies had been collected from 24 FSHD individuals (23 FSHD1 and 1 FSHD2 patient) and 11 controls for RNA-sequencing. From each FSHD patient, two paired MRI-guided [111] muscle biopsies had been collected: one targeting a TIRM⁻ (non-inflamed) muscle (23/24 vastus lateralis, 1/24 vastus intermedius), and one from a muscle with presumed active inflammation selected based on the presence of TIRM hyperintensity, degree of fatty infiltration, and amount and location of normal appearing muscle (15/24 gastrocnemius). Bergström

needle biopsies were collected from the vastus lateralis of all 11 control individuals as previously described [112].

RNA-sequencing

RNA extraction, purification, quality control, library preparation, and sequencing was described previously [60]. Salient points are for PBMCs, globin depletion was performed prior to RNA extraction. Sequencing depth was 21.7–35.5 million reads/sample for muscle samples, and 19.7–46.5 million reads/sample for PBMCs. Raw reads were trimmed with TrimGalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), using CutAdapt [113] to remove Illumina adapters at the 3' end. An additional 15 bases were trimmed from the 5' end of reads due to biased per base sequence content. Human genome sequence GRCh38 and v103 gene annotations were downloaded from Ensembl. Reads were mapped to the human transcriptome using Salmon [114] (v1.6.0), making use of the GC bias correction flag. The RNA-sequencing data from the isogenic FSHD muscle biopsies and PBMCs, and controls is described in Banerji *et al.* [60] and available upon reasonable request from The European Genome-phenome Archive (<https://ega-archive.org>) under accession number EGAS00001007350.

Differential gene expression analysis and clustering

The DESeq2 package [65] in R was used to aggregate length-adjusted transcript counts to the gene-level and perform differential gene expression analyses. A multivariate analysis, adjusting for age and sex, was used to identify differentially expressed genes (DEGs) between control vs. TIRM⁻ muscle, control vs. TIRM⁺ muscle, and control vs. FSHD PBMCs, adjusting for multiple comparisons. To identify genes whose expression differed between TIRM⁻ vs. TIRM⁺ samples, a paired analysis was performed between matched samples from individual patients, adjusting for multiple comparisons. Genes with adjusted *P*-value < 0.05 in each analysis were considered significant.

K-medoids clustering was performed using the ConsensusClusterPlus algorithm [66] employing a Pearson correlation distance metric. For muscle biopsies, the clustering feature space comprised all genes differentially expressed in TIRM⁻ vs control samples and the 500 most significant (as assessed by adjusted *P*-value) genes up- and down-regulated in TIRM⁺ versus control samples, resulting in 1015 genes. For PBMCs, the feature space comprised all genes differentially expressed between control and FSHD PBMCs which were detectable in more than 3 samples, a feature space of 16 genes. The optimal number of clusters was assessed via consensus CDF plots.

Clinical variables (Ricci score, Lamperti score, MRC sum score, LLFF, disease duration, MVC of TA and D4Z4 repeat length) associated with TIRM⁻ FSHD samples assigned to the largest FSHD cluster were compared to those associated with TIRM⁻ FSHD samples assigned to other clusters via multivariate regression, adjusting for age, sex and muscle type. All but one of the TIRM⁻ FSHD samples are vastus lateralis, with the remaining sample vastus intermedius. MVC of TA could not be assessed in the single vastus intermedius sample due to weakness, and thus muscle type was not adjusted for in the regression performed against MVC of TA.

Principal component analysis was performed across muscle biopsy samples and PBMCs, employing the same feature space as for clustering. To confirm the main component of variability in our feature space associated with FSHD/TIRM status, multivariate regression analysis was performed to determine association

between the dominant principal component and FSHD/TIRM status, adjusting for age and sex.

GSEA [115] was performed using Fisher's Exact Test assessing overlap with gene sets compiled by the Molecular Signatures Database (MSigDB) [115]. The following gene sets were investigated: all genes up/down-regulated in TIRM⁻ vs control, the 500 most significant genes up/down-regulated in TIRM⁺ vs control, the 500 most significant genes up/down-regulated in TIRM⁻ vs TIRM⁺, the 16 genes up-regulated across all 3 comparisons (TIRM⁻ vs control, TIRM⁺ vs control and TIRM⁺ vs TIRM⁻), and all genes up/down-regulated in FSHD PBMCs vs controls, separately.

Deconvolution was performed as described by Rubenstein *et al.* [67] using the PLIER package [68] in R to derive latent variables associated with cell types identified in the single cell analysis of human muscle biopsies performed by Rubenstein *et al.* [67]. A single latent variable associated with FAPs but not with any other skeletal muscle cell type, and was assigned to represent the proportion of FAPs across the samples. Multivariate regression analysis was performed to assess association between the FAP associated latent variable and control vs TIRM⁻/TIRM⁺ status adjusting for age and sex. A separate multivariate regression was performed to compare matched TIRM⁻ and TIRM⁺ FSHD samples, adjusting for matched pairs. Significance was assessed at the 5% level.

Expression values of the 13 protein coding mitochondrial genes were summed over for each sample and compared as above via multivariate regression analyses.

Meta-analysis of FAP contributions to FSHD muscle biopsies

Normalized transcriptomic data corresponding to FSHD and control muscle biopsies in seven independent studies [14, 46, 49, 51, 56, 116, 117] was obtained from the GEO database [118]. Rahimov *et al.* [117] GSE36398 describes 50 muscle biopsies assessed by microarray. Bakay *et al.* [116] GSE3307 describes 30 muscle biopsies assessed by microarray. Tasca *et al.* [51] GSE26852 describes 15 muscle biopsies assessed by microarray. Osborne *et al.* [46] GSE10760 describes 49 muscle biopsies assessed by microarray. Dixit *et al.* [14] GSE9397, describes 18 muscle biopsies assessed by microarray. Yao *et al.* [49] GSE56787 describes 23 muscle biopsies assessed by RNA-seq. Wong *et al.* [56] GSE115650 describes 43 muscle biopsies assessed by RNA-seq. Together, these seven datasets describe 228 muscle biopsies (130 FSHD, 98 control).

Each dataset was analyzed independently employing the deconvolution analysis outlined by Rubenstein *et al.* 2020 [67], using the PLIER package [68] in R. For each dataset, a latent variable was identified which associated with FAP cells but not with any other muscle cell type. A two-tailed Wilcoxon test was performed to assess the difference in FAP latent variable values assigned to FSHD vs control samples, with significance assessed at the 5% level. Meta-analysis across the seven independent studies were performed using a random effects model, and overall significance assessed via Fisher's combined test.

Transcript-level analyses

Differential transcript analyses were performed using DRIMSeq [72] and stageR [73] packages in R, using an established protocol [74]. Salmon transcript counts were imported to R using the scaled TPM option in the tximport package [119], a filter was applied to retain transcripts with a read count of at least 10 in all samples and with a relative abundance proportion of less than 0.1 in at least 10 samples. After filtering, DRIMSeq was employed to fit

Dirichlet multinomial models to transcripts, identifying transcript distributions which differed between control, TIRM⁻ and TIRM⁺ muscle or between FSHD and control PBMCs, adjusting for age and sex and multiple testing. We identified genes with evidence of differential transcript usage (DTU) as those with adjusted *P*-values < 0.05. DRIMSeq also evaluated negative binomial models of differential transcript expression (DTE) between control, TIRM⁻ and TIRM⁺ muscle or between FSHD and control PBMCs, within genes with evidence of DTU, adjusting for age, sex and multiple hypotheses. Transcripts with adjusted *P*-values < 0.05 show evidence of DTE. Because DTE results from DRIMSeq have been shown to display an elevated overall false discovery rate (FDR) [74], we employed the *stageR* [73] package in R to identify significant transcripts for which to the overall FDR < 0.05. This refined set of transcripts were considered to show DTE.

GSEA was performed as above, investigating the genes sets containing genes with evidence for both DTU and DTE in TIRM⁻ vs control muscle, TIRM⁺ vs control muscle and FSHD vs control PBMCs.

K-medoids clustering was performed on the PBMC samples as above using the *ConsensusClusterPlus* algorithm [66] employing a Pearson correlation distance metric with a feature space comprising the expression of 40 transcripts with evidence of DTU and DTE between FSHD and control PBMCs.

A data visualization tool for genes which demonstrated evidence of DTU and DTE across our control, TIRM⁻ and TIRM⁺ muscle biopsies and control and FSHD PBMCs was developed using *shiny R* [78]. We used the *ggtranscript* package [120] to visualize transcript structure and the *DRIMSeq* package [72] to visualize DTU and DTE.

Codes written for analysis and visualization of data are available from the corresponding authors upon reasonable request.

Real-time quantitative PCR

RNA was isolated from muscle samples by GeneWiz (<https://www.genewiz.com/en-GB/>), and reverse transcribed using QuantiTect Reverse Transcription kit (Qiagen 205313) as per manufacturer's instructions. RT-qPCR was carried out using Takyon Low ROX SYBR 2X MasterMix blue dTTP (Takyon) as per the manufacturer's instructions on a ViiA7 thermal cycler (Applied Biosystems). TNNT1 ENST00000291901 (containing exon 5) expression was normalized to a region shared between ENST00000291901 and ENST00000356783 (the other isoform showing signs of differential usage), and SAMD4A ENST00000392067 (encoding full-length protein) was normalized relative to ENST00000555719 (truncated isoform). Expression values are presented as $\Delta\Delta C_t$ relative to control. Primer sequences used were:

SAMD4A

ENST00000392067 Forward: 5'-CAGCAGCCAGAATTCCAGC-3' ENST00000555719 Forward: 5'-CAGACCTTAGAGAATTCCAGCTTCC-3' Reverse: 5'-GCGGTCCTCTCGTCTTCAG-3'

TNNT1

ENST00000291901 Forward: 5'-AGGAGGAAGCCCCGAAG-3' ENST00000291901/ENST00000356783 Forward: 5'-GAAGAGGAGGCTCGCGAG-3' Reverse: 5'-CTTCTGGGATCTTTGGCGGG-3'

Acknowledgements

We would like to thank the patients and individuals who generously donated samples.

Supplementary data

Supplementary data is available at *HMG Journal* online.

Conflict of interest statement: BGMvE declares personal fees and non-financial support from Fulcrum, and Facio during the conduct of the study and has a patent EP20120740236 with royalties paid to Euroimmun. ENE, AG, LABJ, PSZ and CRSB declare no conflicts of interest.

Funding

E.N.E. was funded by Wellcome Trust PhD Studentship (WT 222352/Z/21/Z). The Zammit laboratory is funded by the Medical Research Council (MR/P023215/1 and MR/S002472/1), the FSHD Society, Friends research grant (FSHS-82013-06) and Association Française contre les Myopathies. A.G. was funded by Prinses Beatrix Fonds (grant 15-23). BGMvE reports grants from Global FSH, Stichting Spieren voor Spieren, Dutch FSHD Foundation. L.A.B.J. is supported by a Competitiveness Operation Program grant of the Romanian Ministry of European Funds (HINT, ID P_37_762; MySMIS 103587). C.R.S.B. was supported by the Turing-Roche Strategic Partnership.

References

1. Deenen JCW, Arnts H, Van Der Maarel SM. *et al.* Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* 2014;**83**:1056–9.
2. Tawil R. Facioscapulohumeral muscular dystrophy. *Handb Clin Neurol* 2018;**148**:541–8.
3. Tawil R, van der Maarel SM, Tapscott SJ. Facioscapulohumeral dystrophy: the path to consensus on pathophysiology. *Skelet Muscle* 2014;**4**:12–5.
4. Rijken NHM, van der Kooij EL, Hendriks JCM. *et al.* Skeletal muscle imaging in facioscapulohumeral muscular dystrophy, pattern and asymmetry of individual muscle involvement. *Neuromuscul Disord* 2014;**24**:1087–96.
5. Statland JM, Sacconi S, Farmakidis C. *et al.* Coats syndrome in facioscapulohumeral dystrophy type 1: frequency and D4Z4 contraction size. *Neurology* 2013;**80**:1247–50.
6. Lutz KL, Holte L, Kliethermes SA. *et al.* Clinical and genetic features of hearing loss in facioscapulohumeral muscular dystrophy. *Neurology* 2013;**81**:1374–7.
7. Banerji CRS, Zammit PS. Pathomechanisms and biomarkers in facioscapulohumeral muscular dystrophy: roles of DUX4 and PAX7. *EMBO Mol Med* 2021;**13**:e13695.
8. Wijmenga C, Hewitt JE, Sandkuijl LA. *et al.* Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* 1992;**2**:26–30.
9. van Deutekom JCT, Wijmenga C, van Tlenhoven EAE. *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**:2037–42.
10. 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**:1287–95.
11. Lemmers RJLF, Wohlgenuth 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**:884–94.
12. Lemmers RJLF, Tawil R, Petek LM. *et al.* Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele

- causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet* 2012;**44**:1370–4.
13. Gabriëls J, Beckers MC, Ding H. *et al.* Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* 1999;**236**:25–32.
 14. 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 U S A* 2007;**104**:18157–62.
 15. Snider L, Geng LN, Lemmers RJLF. *et al.* Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet* 2010;**6**:e1001181.
 16. Lemmers RJLF, van der Vliet PJ, Klooster R. *et al.* A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 2010;**329**:1650–3.
 17. Greco A, Goossens R, van Engelen B. *et al.* Consequences of epigenetic derepression in facioscapulohumeral muscular dystrophy. *Clin Genet* 2020;**97**:799–814.
 18. Hendrickson PG, Doráis JA, Grow EJ. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet* 2017;**49**:925–34.
 19. Himeda CL, Jones PL. The genetics and epigenetics of facioscapulohumeral muscular dystrophy. *Annu Rev Genomics Hum Genet* 2019;**20**:265–91.
 20. van den Heuvel A, Mahfouz A, Kloet SL. *et al.* Single-cell RNA sequencing in facioscapulohumeral muscular dystrophy disease etiology and development. *Hum Mol Genet* 2019;**28**:1064–75.
 21. Turki A, Hayot M, Carnac G. *et al.* Functional muscle impairment in facioscapulohumeral muscular dystrophy is correlated with oxidative stress and mitochondrial dysfunction. *Free Radic Biol Med* 2012;**53**:1068–79.
 22. Dmitriev P, Bou Saada Y, Dib C. *et al.* DUX4-induced constitutive DNA damage and oxidative stress contribute to aberrant differentiation of myoblasts from FSHD patients. *Free Radic Biol Med* 2016;**99**:244–58.
 23. Knopp P, Krom YD, Banerji CRS. *et al.* DUX4 induces a transcriptome more characteristic of a less-differentiated cell state and inhibits myogenesis. *J Cell Sci* 2016;**129**:3816–31.
 24. Kowaljow V, Marcowycz A, Anseau E. *et al.* The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul Disord* 2007;**17**:611–23.
 25. Wallace LM, Garwick SE, Mei W. *et al.* DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann Neurol* 2011;**69**:540–52.
 26. Banerji CRS, Panamarova M, Hebaishi H. *et al.* PAX7 target genes are globally repressed in facioscapulohumeral muscular dystrophy skeletal muscle. *Nat Commun* 2017;**8**:2152.
 27. Giesige CR, Wallace LM, Heller KN. *et al.* AAV-mediated follistatin gene therapy improves functional outcomes in the TIC-DUX4 mouse model of FSHD. *JCI Insight* 2018;**3**:e123538.
 28. Jones T, Jones PL. A cre-inducible DUX4 transgenic mouse model for investigating facioscapulohumeral muscular dystrophy. *PLoS One* 2018;**13**:e0192657.
 29. Bosnakovski D, Chan SSK, Recht OO. *et al.* Muscle pathology from stochastic low level DUX4 expression in an FSHD mouse model. *Nat Commun* 2017;**8**:1–9.
 30. Pakula A, Lek A, Widrick J. *et al.* Transgenic zebrafish model of DUX4 misexpression reveals a developmental role in FSHD pathogenesis. *Hum Mol Genet* 2019;**28**:320–31.
 31. Jones TI, Chew GL, Barraza-Flores P. *et al.* Transgenic mice expressing tunable levels of DUX4 develop characteristic facioscapulohumeral muscular dystrophy-like pathophysiology ranging in severity. *Skelet Muscle* 2020;**10**:8.
 32. Kan HE, Klomp DWJ, Wohlgemuth M. *et al.* Only fat infiltrated muscles in resting lower leg of FSHD patients show disturbed energy metabolism. *NMR Biomed* 2010;**23**:563–8.
 33. Laoudj-Chenivesse D, Carnac G, Bisbal C. *et al.* Increased levels of adenine nucleotide translocator 1 protein and response to oxidative stress are early events in facioscapulohumeral muscular dystrophy muscle. *J Mol Med* 2005;**83**:216–24.
 34. Heher P, Ganassi M, Weidinger A. *et al.* Interplay between mitochondrial reactive oxygen species, oxidative stress and hypoxic adaptation in facioscapulohumeral muscular dystrophy: metabolic stress as potential therapeutic target. *Redox Biol* 2022;**51**:102251.
 35. Banerji CRS, Zammit PS. PAX7 target gene repression is a superior FSHD biomarker than DUX4 target gene activation, associating with pathological severity and identifying FSHD at the single-cell level. *Hum Mol Genet* 2019;**28**:2224–36.
 36. Banerji CRS. PAX7 target gene repression associates with FSHD progression and pathology over 1 year. *Hum Mol Genet* 2020;**29**:2124–33.
 37. Choi MH, Ow JR, Di Yang N. *et al.* Oxidative stress-mediated skeletal muscle degeneration: molecules, mechanisms, and therapies. *Oxidative Med Cell Longev* 2016;**2016**:6842568.
 38. Guo CY, Sun L, Chen XP. *et al.* Oxidative stress, mitochondrial damage and neurodegenerative diseases. *Neural Regen Res* 2013;**8**:2003–14.
 39. Macaione V, Aguenouz M, Rodolico C. *et al.* RAGE-NF- κ B pathway activation in response to oxidative stress in facioscapulohumeral muscular dystrophy. *Acta Neurol Scand* 2007;**115**:115–21.
 40. Denny AP, Heather AK. Are antioxidants a potential therapy for FSHD? A review of the literature. *Oxidative Med Cell Longev* 2017;**2017**:7020295.
 41. Rickard AM, Petek LM, Miller DG. Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum Mol Genet* 2015;**24**:5901–14.
 42. Jagannathan S, Ogata Y, Gafken PR. *et al.* Quantitative proteomics reveals key roles for post-transcriptional gene regulation in the molecular pathology of facioscapulohumeral muscular dystrophy. *eLife* 2019;**8**:1–15.
 43. Campbell AE, Dyle MC, Matheny T. *et al.* Compromised nonsense-mediated RNA decay results in truncated RNA-binding protein production upon DUX4 expression. *Cell Rep* 2023;**42**:112642.
 44. Banerji CRSS, Knopp P, Moyle LA. *et al.* β -Catenin is central to DUX4-driven network rewiring in facioscapulohumeral muscular dystrophy. *J R Soc Interface* 2015;**12**:20140797.
 45. Ganassi M, Figeac N, Reynaud M. *et al.* Antagonism between DUX4 and DUX4c highlights a pathomechanism operating through β -catenin in facioscapulohumeral muscular dystrophy. *Front Cell Dev Biol* 2022;**10**:802573.
 46. Osborne RJ, Welle S, Venance SL. *et al.* Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. *Neurology* 2007;**68**:569–77.
 47. Winokur ST, Barrett K, Martin JH. *et al.* Facioscapulohumeral muscular dystrophy (FSHD) myoblasts demonstrate increased susceptibility to oxidative stress. *Neuromuscul Disord* 2003;**13**:322–33.

48. Banerji CRS, Panamarova M, Zammit PS. DUX4 expressing immortalized FSHD lymphoblastoid cells express genes elevated in FSHD muscle biopsies, correlating with the early stages of inflammation. *Hum Mol Genet* 2020;**29**:2285–99.
49. Yao Z, Snider L, Balog J. et al. DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum Mol Genet* 2014;**23**:5342–52.
50. Dahlqvist JR, Poulsen NS, Østergaard ST. et al. Evaluation of inflammatory lesions over 2 years in facioscapulohumeral muscular dystrophy. *Neurology* 2020;**95**:e1211–21.
51. Tasca G, Pescatori M, Monforte M. et al. Different molecular signatures in magnetic resonance imaging-staged facioscapulohumeral muscular dystrophy muscles. *PLoS One* 2012;**7**:e38779.
52. Wang LH, Friedman SD, Shaw D. et al. MRI-informed muscle biopsies correlate MRI with pathology and DUX4 target gene expression in FSHD. *Hum Mol Genet* 2019;**28**:476–86.
53. Di Pietro L, Giacalone F, Ragozzino E. et al. Non-myogenic mesenchymal cells contribute to muscle degeneration in facioscapulohumeral muscular dystrophy patients. *Cell Death Dis* 2022;**13**:793.
54. Banerji CRSS, Henderson D, Tawil RN. et al. Skeletal muscle regeneration in facioscapulohumeral muscular dystrophy is correlated with pathological severity. *Hum Mol Genet* 2020;**29**:2746–60.
55. van den Heuvel A, Lassche S, Mul K. et al. Facioscapulohumeral dystrophy transcriptome signatures correlate with different stages of disease and are marked by different MRI biomarkers. *Sci Rep* 2022;**12**:1426.
56. Wong CJ, Wang LH, Friedman SD. et al. Longitudinal measures of RNA expression and disease activity in FSHD muscle biopsies. *Hum Mol Genet* 2020;**29**:1030–43.
57. Bosnakovski D, Xu Z, Ji Gang E. et al. An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J* 2008;**27**:2766–79.
58. Bosnakovski D, Shams AS, Yuan C. et al. Transcriptional and cytopathological hallmarks of FSHD in chronic DUX4-expressing mice. *J Clin Invest* 2020;**130**:2465–77.
59. Molina T, Fabre P, Dumont NA. Fibro-adipogenic progenitors in skeletal muscle homeostasis, regeneration and diseases. *Open Biol* 2021;**11**:210110.
60. Banerji CRS, Greco A, Joosten LAB. et al. The FSHD muscle-blood biomarker: a circulating transcriptomic biomarker for clinical severity in facioscapulohumeral muscular dystrophy. *Brain Commun* 2023;**5**:fcad221.
61. Gros M, Nunes AM, Daoudlarian D. et al. Identification of serum interleukin 6 levels as a disease severity biomarker in facioscapulohumeral muscular dystrophy. *J Neuromuscul Dis* 2022;**9**:83–93.
62. Signorelli M, Mason AG, Mul K. et al. Evaluation of blood gene expression levels in facioscapulohumeral muscular dystrophy patients. *Sci Rep* 2020;**10**:17547.
63. Ricci G, Ruggiero L, Vercelli L. et al. A novel clinical tool to classify facioscapulohumeral muscular dystrophy phenotypes. *J Neurol* 2016;**263**:1204–14.
64. Lamperti C, Fabbri G, Vercelli L. et al. A standardized clinical evaluation of patients affected by facioscapulohumeral muscular dystrophy: the FSHD clinical score. *Muscle Nerve* 2010;**42**:213–7.
65. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550.
66. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010;**26**:1572–3.
67. Rubenstein AB, Smith GR, Raue U. et al. Single-cell transcriptional profiles in human skeletal muscle. *Sci Rep* 2020;**10**:229.
68. Mao W, Zaslavsky E, Hartmann BM. et al. Pathway-level information extractor (PLIER) for gene expression data. *Nat Methods* 2019;**16**:607–10.
69. Banerji CRS, Panamarova M, Pruller J. et al. Dynamic transcriptomic analysis reveals suppression of PGC1 α / ERR α drives muscle atrophy in FSHD. *Hum Mol Genet* 2019;**28**:1244–59.
70. D'Erchia AM, Atlante A, Gadaleta G. et al. Tissue-specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass and respiratory activity. *Mitochondrion* 2015;**20**:13–21.
71. Reznik ED, Wang Q, La K. et al. Mitochondrial respiratory gene expression is suppressed in many cancers. *eLife* 2017;**6**:e21592.
72. Robinson MD, Nowicka M. DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in genomics. *F1000Res* 2016;**5**:1356.
73. Van den Berge K, Sonesson C, Robinson MD. et al. stageR: a general stage-wise method for controlling the gene-level false discovery rate in differential expression and differential transcript usage. *Genome Biol* 2017;**18**:151.
74. Love MI, Sonesson C, Patro R. et al. Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification. *F1000Res* 2018;**7**:952.
75. Contreras O, Rossi FMV, Theret M. Origins, potency, and heterogeneity of skeletal muscle fibro-adipogenic progenitors—time for new definitions. *Skelet Muscle* 2021;**11**:1–25.
76. Careccia G, Saclier M, Tirone M. et al. Rebalancing expression of HMGB1 redox isoforms to counteract muscular dystrophy. *Sci Transl Med* 2021;**13**:eaay8416.
77. Riuzzi F, Sorci G, Sagheddu R. et al. HMGB1–RAGE regulates muscle satellite cell homeostasis through p38-MAPK- and myogenin-dependent repression of Pax7 transcription. *J Cell Sci* 2012;**125**:1440–54.
78. Chang W, Cheng J, Allaire J. et al. Shiny: Web Application Framework for R. R package version 1.2.0. <https://CRAN.R-project.org/package=shiny>.
79. Tawil R, Storvick D, Feasby TE. et al. Extreme variability of expression in monozygotic twins with FSH muscular dystrophy. *Neurology* 1993;**43**:345–8.
80. Lin M-L, Patel H, Remenyi J. et al. Expression profiling of nuclear receptors in breast cancer identifies TLX as a mediator of growth and invasion in triple-negative breast cancer. *Oncotarget* 2015;**6**:21685–703.
81. Curtis C, Shah SP, Chin SF. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;**486**:346–52.
82. van der Maarel M, Frants RR. The D4Z4 repeat-mediated pathogenesis of facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 2005;**76**:375–86.
83. Zatz M, Marie SK, Cerqueira A. et al. The facioscapulohumeral muscular dystrophy (FSHD1) gene affects males more severely and more frequently than females. *Am J Med Genet* 1998;**77**:155–61.
84. Banerji CRS, Cammish P, Evangelista T. et al. Facioscapulohumeral muscular dystrophy 1 patients participating in the UK FSHD registry can be subdivided into 4 patterns of self-reported symptoms. *Neuromuscul Disord* 2020;**30**:315–28.

85. Hogarth MW, Defour A, Lazarski C. *et al.* Fibroadipogenic progenitors are responsible for muscle loss in limb girdle muscular dystrophy 2B. *Nat Commun* 2019;**10**:1–13.
86. Gonzalez D, Contreras O, Rebolledo DL. *et al.* ALS skeletal muscle shows enhanced TGF- β signaling, fibrosis and induction of fibro/adipogenic progenitor markers. *PLoS One* 2017;**12**:e0177649.
87. Bosnakovski D, Oyler D, Mitanoska A. *et al.* Persistent fibroadipogenic progenitor expansion following transient DUX4 expression provokes a profibrotic state in a mouse model for FSHD. *Int J Mol Sci* 2022;**23**:1983.
88. Abbassi-Daloui T, el Abdellaoui S, Voortman LM. *et al.* A transcriptome atlas of leg muscles from healthy human volunteers reveals molecular and cellular signatures associated with muscle location. *eLife* 2023;**12**:e80500.
89. Kendal AR, Layton T, Al-Mossawi H. *et al.* Multi-omic single cell analysis resolves novel stromal cell populations in healthy and diseased human tendon. *Sci Rep* 2020;**10**:1–14.
90. Feng Q, Snider L, Jagannathan S. *et al.* A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. *eLife* 2015;**4**:e04996.
91. Reilich P, Schramm N, Schoser B. *et al.* Facioscapulohumeral muscular dystrophy presenting with unusual phenotypes and atypical morphological features of vacuolar myopathy. *J Neurol* 2010;**257**:1108–18.
92. Cohen J, Huang S, Koczwara K. *et al.* Flavones provide resistance to DUX4-induced toxicity via an mTor-independent mechanism. *Res Sq* [Preprint]. 2023;**2**:rs.3.rs-2452222.
93. La Kethulle D, De Ryhove L, Ansseau E. *et al.* The role of D4Z4-encoded proteins in the osteogenic differentiation of mesenchymal stromal cells isolated from bone marrow. *Stem Cells Dev* 2015;**24**:2674–86.
94. Starkova TY, Polyanichko AM, Artamonova TO. *et al.* Structural characteristics of high-mobility group proteins HMGB1 and HMGB2 and their interaction with DNA. *Int J Mol Sci* 2023;**24**:3577.
95. 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**:339–48.
96. Rinaldi F, Terracciano C, Pisani V. *et al.* Aberrant splicing and expression of the non muscle myosin heavy-chain gene MYH14 in DM1 muscle tissues. *Neurobiol Dis* 2012;**45**:264–71.
97. Kim BJ, Kim AR, Han JH. *et al.* Discovery of MYH14 as an important and unique deafness gene causing prelingually severe autosomal dominant nonsyndromic hearing loss. *J Gene Med* 2017;**19**. <https://doi.org/10.1002/jgm.2950>.
98. Almutawa W, Smith C, Sabouny R. *et al.* The R941L mutation in MYH14 disrupts mitochondrial fission and associates with peripheral neuropathy. *EBioMedicine* 2019;**45**:379–92.
99. Larsson L, Wang X, Yu F. *et al.* Adaptation by alternative RNA splicing of slow troponin T isoforms in type 1 but not type 2 Charcot-Marie-Tooth disease. *Am J Physiol Cell Physiol* 2008;**295**:C722–31.
100. Zhang T, Choi SJ, Wang ZM. *et al.* Human slow troponin T (TNNT1) pre-mRNA alternative splicing is an indicator of skeletal muscle response to resistance exercise in older adults. *J Gerontol A Biol Sci Med Sci* 2014;**69**:1437–47.
101. de Haro M, Al-Ramahi I, Jones KR. *et al.* Smaug/SAMD4A restores translational activity of CUGBP1 and suppresses CUG-induced myopathy. *PLoS Genet* 2013;**9**:e1003445.
102. Chen Z, Holland W, Shelton JM. *et al.* Mutation of mouse Samd4 causes leanness, myopathy, uncoupled mitochondrial respiration, and dysregulated mTORC1 signaling. *Proc Natl Acad Sci U S A* 2014;**111**:7367–72.
103. Kern I, Steimle V, Siegrist C. *et al.* The two novel MHC class II transactivators RFX5 and CIITA both control expression of HLA-DM genes. *Int Immunol* 1995;**7**:1295–9.
104. Villard J, Peretti M, Masternak K. *et al.* A functionally essential domain of RFX5 mediates activation of major histocompatibility complex class II promoters by promoting cooperative binding between RFX and NF- κ B. *Mol Cell Biol* 2000;**20**:3364–76.
105. Pineda JMB, Bradley RK. DUX4 is a common driver of immune evasion and immunotherapy failure in metastatic cancers. *eLife* 2023;**12**:RP89017.
106. Chew GL, Campbell AE, De Neef E. *et al.* DUX4 suppresses MHC class I to promote cancer immune evasion and resistance to checkpoint blockade. *Dev Cell* 2019;**50**:658–671.e7.
107. Jones TI, Himeda CL, Perez DP. *et al.* Large family cohorts of lymphoblastoid cells provide a new cellular model for investigating facioscapulohumeral muscular dystrophy. *Neuromuscul Disord* 2017;**27**:221–38.
108. Kleyweg RP, van der Meché FGA, Schmitz PIM. Interobserver agreement in the assessment of muscle strength and functional abilities in Guillain-Barré syndrome. *Muscle Nerve* 1991;**14**:1103–9.
109. de Jong L, Nikolaev A, Greco A. *et al.* Three-dimensional quantitative muscle ultrasound in a healthy population. *Muscle Nerve* 2021;**64**:199–205.
110. Mul K, Vincenten SCC, Voermans NC. *et al.* Adding quantitative muscle MRI to the FSHD clinical trial toolbox. *Neurology* 2017;**89**:2057–65.
111. Lassche S, Janssen BH, Ijzermans T. *et al.* MRI-guided biopsy as a tool for diagnosis and research of muscle disorders. *J Neuromuscul Dis* 2018;**5**:315–9.
112. Tarnopolsky MA, Pearce E, Smith K. *et al.* Suction-modified Bergström muscle biopsy technique: experience with 13,500 procedures. *Muscle Nerve* 2011;**43**:716–25.
113. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011;**17**:10–2.
114. Patro R, Duggal G, Love MI. *et al.* Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 2017;**14**:417–9.
115. Subramanian A, Tamayo P, Mootha VK. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;**102**:15545–50.
116. Bakay M, Wang Z, Melcon G. *et al.* Nuclear envelope dystrophies show a transcriptional fingerprint suggesting disruption of Rb-MyoD pathways in muscle regeneration. *Brain* 2006;**129**:996–1013.
117. Rahimov F, King OD, Leung DG. *et al.* Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers. *Proc Natl Acad Sci U S A* 2012;**109**:16234–9.
118. Barrett T, Wilhite SE, Ledoux P. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2012;**41**:D991–5.
119. Sonesson C, Love MI, Robinson MD. *et al.* Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 2016;**4**:1521.
120. Gustavsson EK, Zhang D, Reynolds RH. *et al.* ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2. *Bioinformatics* 2022;**38**:3844–6.