Tracking disease progression non-invasively in Duchenne and Becker muscular dystrophies

Pietro Spitali1*, Kristina Hettne1, Roula Tsonaka2, Mohammed Charrouf1, Janneke van den Bergen3, Zaida Koeks3, Hermien E. Kan4, Melissa T. Hooijmans4, Andreas Roos5, Volker Straub5, Francesco Muntoni6, Cristina Al-Khalili-Szigyarto7, Marleen J.A. Koel-Simmelink8, Charlotte E. Teunissen8, Hanns Lochmüller5, Erik H. Niks3 & Annemieke Aartsma-Rus1,4

1Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands, 2Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands, 3Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands, 4C.J. Gorter Center for High Field MRI, Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands, 5John Walton Muscular Dystrophy Research Centre, MRC Centre for Neuromuscular Diseases, Institute of Genetic Medicine, University of Newcastle, Newcastle upon Tyne, UK, 6Dubowitz Neuromuscular Centre, University College London Great Ormond Street Institute of Child Health, London, UK, 7Division of Proteomics, School of Biotechnology, KTH—Royal Institute of Technology, Stockholm, Sweden, 8Neurochemistry Lab and Biobank, Department of Clinical Chemistry, Amsterdam Neuroscience, VU University Medical Center Amsterdam, The Netherlands

Abstract

Background Analysis of muscle biopsies allowed to characterize the pathophysiological changes of Duchenne and Becker muscular dystrophies (D/BMD) leading to the clinical phenotype. Muscle tissue is often investigated during interventional dose finding studies to show in situ proof of concept and pharmacodynamics effect of the tested drug. Less invasive readouts are needed to objectively monitor patients’ health status, muscle quality, and response to treatment. The identification of serum biomarkers correlating with clinical function and able to anticipate functional scales is particularly needed for personalized patient management and to support drug development programs.

Methods A large-scale proteomic approach was used to identify serum biomarkers describing pathophysiological changes (e.g. loss of muscle mass), association with clinical function, prediction of disease milestones, association with in vivo 31P magnetic resonance spectroscopy data and dystrophin levels in muscles. Cross-sectional comparisons were performed to compare DMD patients, BMD patients, and healthy controls. A group of DMD patients was followed up for a median of 4.4 years to allow monitoring of individual disease trajectories based on yearly visits.

Results Cross-sectional comparison enabled to identify 10 proteins discriminating between healthy controls, DMD and BMD patients. Several proteins (285) were able to separate DMD from healthy, while 121 proteins differentiated between BMD and DMD; only 13 proteins separated BMD and healthy individuals. The concentration of specific proteins in serum was significantly associated with patients’ performance (e.g. BMP6 serum levels and elbow flexion) or dystrophin levels (e.g. TIMP2) in BMD patients. Analysis of longitudinal trajectories allowed to identify 427 proteins affected over time indicating loss of muscle mass, replacement of muscle by adipose tissue, and cardiac involvement. Over-representation analysis of longitudinal data allowed to highlight proteins that could be used as pharmacodynamic biomarkers for drugs currently in clinical development.

Conclusions Serum proteomic analysis allowed to not only discriminate among DMD, BMD, and healthy subjects, but it enabled to detect significant associations with clinical function, dystrophin levels, and disease progression.

Keywords Biomarker; Duchenne muscular dystrophy; Becker muscular dystrophy; Longitudinal analysis; Proteomics; Sarcopenia

Received: 14 November 2017; Revised: 25 January 2018; Accepted: 10 March 2018

*Correspondence to: Pietro Spitali, Department of Human Genetics, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands. Tel: +31 715269437, Fax: +31 715268285, Email: p.spitali@lumc.nl

© 2018 The Authors. Journal of Cachexia, Sarcopenia and Muscle published by John Wiley & Sons Ltd on behalf of the Society on Sarcopenia, Cachexia and Wasting Disorders. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
Introduction

Duchenne muscular dystrophy (DMD) is a severe muscle-wasting disease caused by protein truncating mutations in the dystrophin encoding DMD gene.\textsuperscript{1–4} The identification of biomarkers for DMD is a priority to the field to identify prognostic factors and to decrease over time as muscle mass is lost.\textsuperscript{15} In several muscle specific molecular entities have been detected in biofluids such as serum, plasma, and urine showing that it is possible to obtain direct muscle-related information without an invasive muscle biopsy.\textsuperscript{6–15} Muscle-derived proteins such as muscle creatine kinase (CK) are known to be highly elevated in patients’ blood in the early phases of the disease and to decrease over time as muscle mass is lost.\textsuperscript{15} In several interventional clinical trials, CK levels dropped, with a range of possible plausible explanations such as reduced muscle damage after treatment, but also reduced exercise due to hospitalization or even a decrease in muscle mass. Indeed, other factors that make interpretation of CK levels difficult are day-to-day variability, seasonal variability, and exercise-dependent fluctuations. All these represent obstacles for the use of CK as pharmacodynamic biomarker.

Other muscle-related proteins such as myosin light chain 3 (MYL3) show profiles comparable to CK and as such probably also make poor pharmacodynamic biomarkers.\textsuperscript{8,11} Only a few proteins were reported to be stably elevated in DMD such as MMP-9\textsuperscript{6} and LEP.\textsuperscript{16} While these may be attractive biomarkers, little is known about whether serum concentration vary after therapeutic intervention in animal studies or clinical trials.

So far, studies in the literature have focused on the identification of biomarkers able to discriminate between healthy and DMD using different technologies, describing how the identified proteins belong to pathways known to be associated with lack of dystrophin or with the secondary pathology. While these studies provide complementary information about the disease status, longitudinal studies allowing to model individual disease trajectories are lacking. This is particularly important given that clinical differences are observed in patients carrying the same or comparable causative mutation. Published studies often used age as a proxy for disease severity,\textsuperscript{8} due to the unavailability of longitudinal samples that would enable modelling the intra-individual variation and that represent a better simulation of evolving pathologic processes. There is only one recent report showing how a restricted number of proteins and steroid hormones can monitor safety and efficacy of glucocorticoids treatment in a longitudinal setup.\textsuperscript{17}

In this study, we assessed protein concentrations in serum samples of DMD patients and healthy controls to identify relevant biological molecules able to discriminate between healthy and disease. We then compared the identified signature with patients affected by the milder allelic form of the disease called Becker muscular dystrophy (BMD). We further examined BMD patients to identify associations between serum proteomic data and clinical data, in vivo \textsuperscript{31}P phosphorous magnetic resonance spectroscopy (\textsuperscript{31}P MRS) data, muscle derived gene expression data, and dystrophin levels in muscle. Finally, we modelled longitudinal proteomic profiles of a cohort of DMD patients followed up for several years, providing estimates of proteomic changes over time and prediction of disease milestones.

Materials and methods

Patients participants

In the cross-sectional comparison, 15 DMD patients and 9 healthy age matched controls were included. The median age was 9 years for both groups, while the mean was 9 years for DMD and 10 years for the control group. Metadata are provided in Table S1. A total of 62 BMD patients were studied of whom 33 were followed up at the Leiden University Medical Center (LUMC) and 29 at the John Walton Muscular Dystrophy Centre of the University of Newcastle. Ten healthy adult controls were included. The median age was 39 (age range 19–66) for the Leiden BMDs, 35 (age range 19–67) for the Newcastle BMDs, and 46 (age range 24–64) for the healthy controls. Muscle strength was measured for a subset of BMD patients by using a quantitative muscle assessment (QMA) system (www.qmasystem.com) to assess the maximal voluntary isometric contraction.\textsuperscript{18} Measurements of the following muscle groups were obtained: handgrip, elbow flexion and extension, and knee and hip flexion and extension. Measurements were performed as described by Hogrel et al.\textsuperscript{19} For each muscle, the highest value of three consecutive measurements was used as the maximum strength (in kg). If a measurement differed more than 10% from previous measurements, it was discarded, and an extra measurement was performed. Handgrip strength was assessed on the dominant hand. Participants were sitting in an upright position with their arm parallel to their trunk and were encouraged to squeeze the dynamometer of the QMA system as hard as possible. Elbow flexion was also assessed in a supine position with the forearm in a neutral position and the elbow at the side flexed at 90° and the strap attached around the wrist. Hip flexion and extension were assessed in supine position with the hip and knee flexed at 90° and the strap attached proximal to the knee. For knee flexion and knee extension, patients were tested in sitting position, with hip and knee flexed at 90° flexion with the strap attached around their ankle. The highest value of three consecutive measurements was used as the maximum strength. If a measurement differed more than 10% from previous measurements, it was discarded, and an extra
measurement was performed. Dystrophin was quantified for 25 patients. $^{31}$P MRS data were available for 23 BMD patients. Twenty-one muscle biopsies were available to perform gene expression analysis.

Fourteen DMD patients were enrolled in a longitudinal study, five of whom were also included in the cross-sectional study. The median age was 8.6 years (range 5.4–14.4 years of age). Eleven of them were followed-up at the LUMC, while three were followed-up at the John Walton Muscular Dystrophy Research Centre in Newcastle. The median follow-up time was 4.4 years (range 0.4–5.3 years). A total of 56 samples were analysed with an average of four samples per patients. All patients’ information is reported in Table S2.

The study has been approved by the Institutional Review Board of the involved clinical centers. Informed consent forms were obtained for all participants. The investigation was conducted according to the declaration of Helsinki.

Serum preparation

Venous blood was allowed to clot for 30 min in red capped tubes followed by centrifugation for 10 min at 2350 g. Serum was carefully removed, aliquoted, and conserved at −80 °C pending use.

Proteomics

The commercially available SOMAscan® multiplex assay consists of individual affinity molecules called SOMAmer® (slow off-rate modified DNA aptamer) reagents, each with very high affinity to their protein targets. In brief, a biological sample in each well of a 96-well plate is incubated with a mixture of the SOMAmer reagents. After bead-based immobilization and washing steps, target-bound SOMAmer reagents were purified and quantified on a custom Agilent hybridization array. For the cross-sectional analysis, a mix of 1305 somamers was used (performed by the Neurochemistry Lab at VU University Medical Center Amsterdam, The Netherlands), while for the longitudinal study a mix of 4006 somamers was used (performed at Somalogic Inc., Boulder, USA).

Analysis of muscle tissue

Gene expression analysis was performed in muscle biopsies obtained for a subset of BMD patients. Healthy control muscle derived total RNA was purchased from Ambion. To obtain gene expression data, total RNA was purified using Tripure and resuspended in RNase-free water. cDNA synthesis was performed using the Transcriptor Reverse Transcriptase (cat. n. 03531287001, Roche), random hexamers (cat. n. SO142, ThermoFisher Scientific), RNAsin ribonuclease inhibitor (cat. n. N2515, Promega), and dNTPs (cat. n. 10297018, ThermoFisher Scientific) in a total volume of 20 μL according to the manufacturer’s instructions. A selection of genes was performed based on genes known to be differentially expressed in DMD based on available human and mouse data. Primer pairs were designed spanning at least one splice junction, and RT-qPCR was used to quantify gene expression using SYBR green as intercalating dye. Melting curve analysis was performed to filter out primers leading to non-specific amplification products. Data analysis was performed using the LinReg software to correct for PCR efficiency. The fold change over healthy skeletal muscle derived total RNA (cat. n. AM7982, Ambion) was used for downstream analysis. Dystrophin quantification was performed by western blot as previously described.

$^{31}$P phosphorous magnetic resonance spectroscopy

In vivo $^{31}$P MRS datasets were obtained from the right lower leg using a 7T MR scanner (Philips, Achieva, Best, The Netherlands) for 23 BMD patients as described before. Metabolite ratios and tissue pH were calculated for muscles individually and then averaged for all muscles. The following ratios were used: inorganic phosphate over ATP (Pi/ATP) Pi over phosphocreatine (Pi/PCr), phosphodiesters over ATP (PDE/ATP), and PCr over ATP (Pcr/ATP).

Statistics and pathway analysis

Analysis of proteomic cross-sectional data was performed after hybridization control normalization, median signal normalization, and log10 transformation. A linear model with FDR multiple testing correction was used to compare protein levels in DMD patients and controls. The analysis of BMD patients focused on those proteins found to be differentially expressed in DMD based on known to be differentially expressed. A selection of genes was performed based on genes known to be differentially expressed in DMD based on available human and mouse data. Primer pairs were designed spanning at least one splice junction, and RT-qPCR was used to quantify gene expression using SYBR green as intercalating dye. Melting curve analysis was performed to filter out primers leading to non-specific amplification products. Data analysis was performed using the LinReg software to correct for PCR efficiency. The fold change over healthy skeletal muscle derived total RNA (cat. n. AM7982, Ambion) was used for downstream analysis. Dystrophin quantification was performed by western blot as previously described.

Analysis of muscle tissue

Gene expression analysis was performed in muscle biopsies obtained for a subset of BMD patients. Healthy control muscle derived total RNA was purchased from Ambion. To obtain gene expression data, total RNA was purified using Tripure and resuspended in RNase-free water. cDNA synthesis was performed using the Transcriptor Reverse Transcriptase (cat. n. 03531287001, Roche), random hexamers (cat. n. SO142, ThermoFisher Scientific), RNAsin ribonuclease inhibitor (cat. n. N2515, Promega), and dNTPs (cat. n. 10297018, ThermoFisher Scientific) in a total volume of 20 μL according to the manufacturer’s instructions. A selection of genes was performed based on genes known to be differentially expressed in DMD based on available human and mouse data. Primer pairs were designed spanning at least one splice junction, and RT-qPCR was used to quantify gene expression using SYBR green as intercalating dye. Melting curve analysis was performed to filter out primers leading to non-specific amplification products. Data analysis was performed using the LinReg software to correct for PCR efficiency. The fold change over healthy skeletal muscle derived total RNA (cat. n. AM7982, Ambion) was used for downstream analysis. Dystrophin quantification was performed by western blot as previously described.

Analysis of muscle tissue

Gene expression analysis was performed in muscle biopsies obtained for a subset of BMD patients. Healthy control muscle derived total RNA was purchased from Ambion. To obtain gene expression data, total RNA was purified using Tripure and resuspended in RNase-free water. cDNA synthesis was performed using the Transcriptor Reverse Transcriptase (cat. n. 03531287001, Roche), random hexamers (cat. n. SO142, ThermoFisher Scientific), RNAsin ribonuclease inhibitor (cat. n. N2515, Promega), and dNTPs (cat. n. 10297018, ThermoFisher Scientific) in a total volume of 20 μL according to the manufacturer’s instructions. A selection of genes was performed based on genes known to be differentially expressed in DMD based on available human and mouse data. Primer pairs were designed spanning at least one splice junction, and RT-qPCR was used to quantify gene expression using SYBR green as intercalating dye. Melting curve analysis was performed to filter out primers leading to non-specific amplification products. Data analysis was performed using the LinReg software to correct for PCR efficiency. The fold change over healthy skeletal muscle derived total RNA (cat. n. AM7982, Ambion) was used for downstream analysis. Dystrophin quantification was performed by western blot as previously described.

Analysis of muscle tissue

Gene expression analysis was performed in muscle biopsies obtained for a subset of BMD patients. Healthy control muscle derived total RNA was purchased from Ambion. To obtain gene expression data, total RNA was purified using Tripure and resuspended in RNase-free water. cDNA synthesis was performed using the Transcriptor Reverse Transcriptase (cat. n. 03531287001, Roche), random hexamers (cat. n. SO142, ThermoFisher Scientific), RNas
convergence problems were observed, we have considered a simpler random effects structure, i.e. a single random intercepts term. We tested if the mean change in time of each protein is statistically significant using the likelihood ratio test. All analyses have been performed in R. Pathway analysis of both cross-sectional and longitudinal proteomic data was performed using the online tool GeneTrail2 (version 1.5).27 The over-representation analysis was applied using as reference gene-set all proteins detected by the SOMAscan assay depleted by the significant ones. Wikipathways was the sole category tested for this analysis.

Results

Cross-sectional analysis of DMD, BMD, and healthy controls

Proteomic analysis of serum samples by SOMAscan enabled the quantification of 1305 proteins in 15 DMD patients and nine healthy controls. A total of 285 aptamers (directed to a total of 276 unique targets) detected significant changes between DMD patients and healthy controls (Figure 1A and Table S3); protein concentration was reduced in DMD patients for 120 proteins, while 165 proteins were found to be elevated in DMD patients. Overlap analysis with two previously published datasets comparing DMD and healthy individuals using the same technology8 revealed that 34 proteins were able to discriminate between DMD patients and healthy controls in all cohorts (Figure 1B). All 34 proteins except GAPDH showed concordant directional change towards either a consistent increase or decrease in patients compared with healthy controls in the two studies (Table 1), suggesting that 33 proteins are bona fide diagnostic biomarkers for DMD. Two of the identified proteins (APOE and C3) were quantified with three independent aptamers, and another four (F9, FGA-FGB-FGG, LYN, and PLG) were detected with two different aptamers. The directional change was consistent for all proteins recognized by multiple SOMAmers indicating a first layer of internal validation for these entities. Examples of down-regulated and up-regulated proteins are provided in Figure 1C. Hierarchical clustering of the differentially expressed proteins identified in healthy and DMD individuals followed by correlation analysis showed that many of the differentially expressed proteins are co-expressed in healthy individuals and that co-expression is lost in the DMD context (Figure 1D-E). Pathway analysis identified 44 pathways affected in DMD patients compared with healthy controls encompassing insulin signalling, energy-related pathways such as glycolysis and gluconeogenesis, Wnt signalling, TGFβ pathway as well as pathways affecting cardiac tissue (e.g. calcium regulation of cardiac cells, microRNAs in cardiomyocyte hypertrophy, and cardiac hypertrophic response) (Table S4).

To understand which proteins could be affected by dystrophin presence or dystrophin reintroduction in DMD patients, we considered the 285 proteins affected in DMD patients and studied whether their concentration was affected in two independent cohorts of BMD patients; 13 of these SOMAmer targets were able to discriminate between BMD patients and healthy individuals (Figure 2A). Four proteins were reduced in BMD patients compared with healthy controls, while nine proteins were found to be elevated. The identified proteins were involved in muscle function and contraction. All markers except 2 (PGD and MDK) showed changes in the same direction as for DMD patients, but the amplitude of the change was somewhat lower in BMD patients compared with DMD (e.g. CA3 and MAPK12) (Table 2). The comparison of DMD and BMD individuals highlighted 121 SOMAmer targets as being differentially represented, with 33 proteins being elevated and 88 being reduced in BMD compared with DMD patients (Figure 2B). Overlap analysis showed that 10 proteins were able to discriminate between the three groups (Table 2).

For the Dutch cohort of BMD patients, 31P MRS data, upper and lower limb functional data, ambulation status, dystrophin levels in muscle, and targeted gene expression data were obtained, allowing to investigate whether associations exist with circulating proteins levels. No significant associations were found between serum protein levels, 31P MRS, and muscle derived gene expression data after multiple testing correction. Serum concentration of four proteins encoded by genes IGF2R, CDNF, RGMA, and BMP6 (corresponding to Uniprot IDs P11717, Q49AH0, Q96B86, and P22004) were associated with upper limb function as measured by elbow flexion (Figure 3A-C). TIMP2 and SERPIND1 (Uniprot IDs P16035 and P05546) were significantly associated with dystrophin levels in muscle quantified by western blot (Figure 3D).

Longitudinal analysis of Duchenne muscular dystrophy patients

To study whether disease progression is reflected in the biomarker peripheral signature over time, we studied a longitudinal cohort composed of 14 patients followed up for an average of 4 years (Table S2). A total of 4006 proteins were quantified in 56 samples. The vast majority were unchanged (3579 constituting 89% of all quantified proteins), while 427 (427 unique Uniprot IDs) changed over time: 83 (2%) increased, while 344 (9%) decreased over time (Figure 4A). The estimated change per year in log scale per protein was calculated and is presented in Figure 4B. Proteins down-regulated over time were enriched for muscle specific proteins or proteins involved in muscle function, while proteins up-regulated over time were mostly proteins involved in fat formation or response to oxidative stress. The proteins with
the largest estimated increase over time were leptin (gene name LEP, estimated mean log increase of 0.289 per year), nicotinamide phosphoribosyltransferase (gene name NAMPT), and carnosine dipeptidase 1 (gene name CNDP1); the proteins with largest estimated decrease over time were phospholipase A2 group IIA (gene name PLA2G2A), myosin binding protein C slow type (gene name MYBPC1), and creatine kinase M-type (gene name CKM). The estimated changes per year for each individual protein are listed in Table S5. Among the 427 proteins affected by time, 66 were differentially represented between cases and controls in the cross-sectional proteomic study, underlining that most of the proteins changing over time and possibly tracking disease progression are not differentially represented between patients and controls (Figure 4C). Among these 66, only six (SNRPF, N6AMT1, ERP29, CNTFR, KEAP1, and C5) increased over time.
in DMD patients; four of these (SNRPF, N6AMT1, ERP29, and KEAP1) were elevated compared with healthy controls. The remaining proteins differentially represented between cases and controls did not change over time constituting a stable set of proteins able to discriminate between healthy and disease status.

Given that all patients were able to walk when the first sample was obtained and wheelchair dependent when the last sample was collected, we investigated whether proteins were associated with an increased risk of wheelchair dependency. None of the 4006 proteins showed a significant association after multiple testing correction; however, 34 proteins had unadjusted $P$-value below 0.05, 18 of which with hazard ratio above 2, and four of them with hazard ratio below 0.5 (Figure 4F and Table S6). The most extreme positive hazard ratio was recorded for CPB1 meaning that a unit increase (in a log2 transformed scale) in CPB1 was associated with a 414% increased risk of wheelchair dependency. Over-representation analysis of the proteins affected over time enabled the identification of 27 pathways affected in DMD patients during disease progression (23 and 4 represented, respectively, by the proteins decreasing and increasing over time) (Table S4). The pathways represented by the proteins reduced over time were typical DMD pathways (e.g. striated muscle contraction, calcium regulation, energy production pathways, and focal adhesion), while adipogenesis, differentiation of white and brown adipocytes, transcriptional regulation of white adipocyte differentiation, and the NRF2 pathway were enriched in subset of proteins increasing over time.

**Discussion**

DMD is a lethal disease caused by the absence of dystrophin, leading to chronic muscle damage followed up by fibro-fatty substitution of muscle mass.\textsuperscript{28,29} Downstream effects of lack of dystrophin have been characterized by histological, gene expression, and proteomic analyses of muscles biopsies obtained from affected patients (and animal models) enabling the identification of morphological alterations and pathological pathways behind the clinical presentation.\textsuperscript{30–33} The examination of muscle biopsies is still often performed during...
interventional clinical trials to study the effect of the drug (e.g. dystrophin restoration by antisense oligonucleotides or read-through compounds or levels of fibrosis by anti-fibrotic compounds). Obtaining muscle biopsies is however an invasive procedure that both patients and investigators would ideally avoid if similar information could be obtained by studying blood derived samples.

In this work, we first studied a large number of proteins in serum to identify blood biomarkers reflecting the pathological changes occurring in striated muscles of patients affected by DMD compared with healthy controls. Our observations show that several of the known pathological pathways known to affect DMD muscles can be monitored peripherally. We confirm many of the already known proteins

---

**Figure 2** Cross-sectional analysis of BMD patients. (A) Volcano plot showing the estimated change in BMD patients compared with healthy controls (x-axis) and the \(-\log_{10}\) of the adjusted \(P\)-values. Black circles represent proteins that are not differentially represented in BMD patients compared to healthy controls, while red circles represent the 13 proteins surviving multiple testing correction. (B) Volcano plot showing the estimated change in BMD patients compared with DMD patients. Red circles represent the 121 proteins that were significant after multiple testing correction. (C) Venn diagram showing the overlap between proteins differentially expressed in the DMD, BMD, and healthy. (D–F) Dot plots showing three examples of proteins able to discriminate among DMD, BMD, and healthy (D), or discriminating between DMD and healthy but not between BMD and healthy (E), or able to discriminate between dystrophic and non-dystrophic but unable to separate DMD and BMD (F).

**Table 2** List of proteins differentially represented in sera of BMD patients compared with healthy controls. The table also shows whether the directional change is in line with the observations in DMD patients.

<table>
<thead>
<tr>
<th>Uniprot ID</th>
<th>Gene name</th>
<th>Fold change</th>
<th>(P)-value</th>
<th>Adjusted (P)-value</th>
<th>Same direction as DMD</th>
<th>Different in DMD vs. BMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>P07451</td>
<td>CA3</td>
<td>5.33</td>
<td>2.11E-10</td>
<td>6.03E-08</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>P48788</td>
<td>TNNI2</td>
<td>3.53</td>
<td>4.70E-06</td>
<td>6.70E-04</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>O95390/O14793</td>
<td>GDF11 MSTN</td>
<td>0.66</td>
<td>8.04E-06</td>
<td>7.64E-04</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>P52209</td>
<td>PGD</td>
<td>0.35</td>
<td>2.68E-05</td>
<td>1.84E-03</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Q00872</td>
<td>MYBPC1</td>
<td>2.13</td>
<td>3.22E-05</td>
<td>1.84E-03</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>P53778</td>
<td>MAPK12</td>
<td>1.48</td>
<td>4.81E-05</td>
<td>2.28E-03</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>P02144</td>
<td>MB</td>
<td>1.97</td>
<td>1.40E-04</td>
<td>4.98E-03</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>P12277</td>
<td>CKB CKM</td>
<td>3.65</td>
<td>1.40E-04</td>
<td>4.98E-03</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>P06732</td>
<td>CKM</td>
<td>2.70</td>
<td>1.63E-04</td>
<td>5.16E-03</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>P24298</td>
<td>GPT</td>
<td>1.39</td>
<td>7.45E-04</td>
<td>2.03E-02</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>P41217</td>
<td>CD200</td>
<td>0.84</td>
<td>7.84E-04</td>
<td>2.03E-02</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>P04075</td>
<td>ALDOA</td>
<td>1.76</td>
<td>9.95E-04</td>
<td>2.36E-02</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>P21741</td>
<td>MDK</td>
<td>0.49</td>
<td>1.39E-03</td>
<td>3.04E-02</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
found to be differentially represented between cases and controls and provides a list of 33 bona fide diagnostic biomarkers able to discriminate between healthy and diseased. We expanded the list with 165 new proteins that were not identified in previous studies. Interestingly, many pathways (e.g. focal adhesion and striated muscle contraction) that have been identified by gene expression studies of muscle biopsies\textsuperscript{31,34} were detected and quantified by the analysis of serum samples, providing evidence that muscle biopsies could be avoided when the aim is to monitor these particular pathways.

We then compared the obtained signature with two independent cohorts of BMD patients. A strong normalization towards healthy controls was observed in BMD; only 13 proteins were found to be affected compared with healthy controls, while 121 proteins were differentially represented in BMD compared with DMD. Interestingly, 10 proteins were able to discriminate between the three groups, indicating that a shift towards the levels observed in BMD patients could be observed following dystrophin restoration in DMD patients.

The availability of in vivo \textsuperscript{31}P MRS data as well as functional data, gene expression data, and dystrophin levels in muscle for the Dutch cohort of BMD patients enabled us to study whether associations exist between these outcomes and serum protein levels. We identified four proteins associated with upper limb function and two proteins associated with dystrophin levels as assessed by Western blot. Three of the four proteins correlating with elbow flexion were related to TGF-\(\beta\): IGF2R is involved in the control of extracellular levels TGF-\(\beta\) (even though not specifically for muscle),\textsuperscript{35} RGMA functions as a bone morphogenetic protein (BMP) co-receptor enhancing BMP signalling,\textsuperscript{36} and BMP6 was discovered to attenuate TGF-\(\beta\) signalling in Dupuytren derived fibroblasts.\textsuperscript{37} The last protein correlating with elbow flexion was CDNF, a neurotrophic factor acting on dopamine

---

**Figure 3** Association of serum proteomic signature with \textsuperscript{31}P MRS data, functional data, and dystrophin levels. (A) Correlation plot showing the high degree of (Pearson) correlation between the different functional scores and moderate correlation among \textsuperscript{31}P MRS data with the exception of the Pi/PCr and Pi/ATP ratios. This plot was used to identify potential correlating covariates to exclude during modelling. (B) Heat map showing the Pearson correlation of IGF2R, CDNF, RGMA, and BMP6 with upper and lower limb functional data. (C) Scatter plot showing the linear association between elbow flexion and the concentration of these four proteins in serum. (D) Scatter plot showing the association between TIMP2 and SERPIND1 with dystrophin protein levels in muscle.
Figure 4 Longitudinal proteomic analysis of DMD patients. (A) Pie chart showing the number of proteins unaffected, increasing, and decreasing over time in DMD patients followed-up longitudinally. (B) Volcano plot showing the estimated mean log change per year (x axis) for each protein and the $-\log_{10}$ of the P-value (y axis). Black circles represent proteins stable during disease progression, green and red circles represent proteins with decreasing and increasing concentration over time, respectively. (C) Venn diagram showing the overlap of proteins identified in the cross-sectional and longitudinal proteomic studies. (D–E) Scatter plots showing exemplar proteins decreasing (D) and increasing (F) over time in DMD patients. Age is plotted on the x axis, while protein concentration is plotted on the y axis. (F) Volcano plot showing the hazard ratios (x axis) and $-\log_{10}$ of the P-value of the survival analysis modelling the time to loss of ambulation. No protein is significant after multiple testing correction.

To further explore whether peripheral proteomic profiling could be used to model disease progression, we studied the concentration of 4006 proteins in DMD patients followed up over time for a median of 4.4 years. Importantly, our study is the first to quantify and analyse a large number of proteins in a longitudinal setting, providing a detailed overview of the pathological features that can be monitored peripherally as disease progresses. In total, 427 proteins were significantly affected by disease progression. The estimated changes of each protein per year were computed. Most of the proteins which decreased over time (344) were related to muscle degeneration, heart condition, focal adhesion, and energy production; the 83 proteins which increased over time were mostly connected to fat formation and oxidative stress. The downtrend of muscle related proteins has been widely described in DMD patients, and it has been related to disease caused muscle damage/breakdown as well as muscle weakness, sarcopenia, and cachexia. It is not clear to which extent the observed changes are caused by the disease itself or to the inability of the patients to use the affected muscles. There are only a few papers reporting the association of blood derived adipose markers such as leptin with DMD. The increase in leptin levels has been associated with fat mass and treatment with corticosteroids in DMD patients; an increase in leptin has however been reported in steroid naïve DMD patients together with other markers of metabolic syndrome such as hypertriglyceridemia, hyperinsulinemia and insulin resistance. While the fat derived signature present in blood is certainly confounded by the treatment with corticosteroids in our study, we did not observe changes characteristic of the metabolic syndrome such as an increase of glucose levels (mean 5.5 standard deviation 1.2 mmol/L of non-fasting glucose) and glycated haemoglobin (mean 37.7 standard deviation 1.9 mmol/mol Hb). Because biomarker profiling shows some signs of pre-symptomatic metabolic syndrome, we argue that the
substitution of muscle mass by adipose tissue (hallmark of the disease) contributes to the fat derived signature observed in patients’ serum with markers such as AGT, CNTFR, LEP, NAMPT, and SOCS3. Prospective studies including both MRI, histological analysis, and serum biomarkers should aim to clarify whether a direct and linear relationship exists between muscle fat fraction and peripheral markers of adipose tissue.

Modelling of longitudinal profiles enabled us to identify proteins that could be used as pharmacodynamic and safety biomarkers in dose finding studies for different types of drugs. Several drugs aim to correct different secondary effects of lack of dystrophin leading to the need of different biomarkers in order to be able to monitor the drug effects. As an example, effects of drugs aiming to increase muscle mass (e.g. IGF-1 overexpression\textsuperscript{33,44}) could be monitored by studying AKT protein concentration in serum, which was decreasing over time in the studied population; a second example could be a drug aiming to stabilize calcium homeostasis in muscle fibres and in cardiomyocytes (e.g. Rycalls), the effect of which may be monitored by e.g. gene ATP1B2\textsuperscript{45} and Ca\textsuperscript{2+}/calmodulin-dependent protein kinases (CAMK2A, CAMK2B, and CAMK2D\textsuperscript{46}) all linked to calcium homeostasis in muscle cells and captured in our assays. Not all therapeutic approaches may however find a direct pharmacodynamic biomarker to be evaluated in serum as dystrophin restoring drugs are likely to still require muscle biopsies to prove correct dystrophin expression and localization in dose optimization trials; however dystrophin restoration should result in an overall normalization of the signal towards healthy controls that could be monitored peripherally and hopefully avoid the need for muscle biopsies in the larger phase III studies.

Our study has some limitations such as the long intervals between samples, the overall longitudinal study duration poorly reflecting the duration of clinical trials, the lack of controls for food intake and corticosteroid use. Another limitation of our study is represented by the impossibility to test whether associations exist between serum biomarkers levels and patients walking performance (e.g. 6-minute walk test data), cardiac and pulmonary data as well as daily activity data as these data were not available or there were too many missing data points. Despite these limitations, our study is the first study modelling a large number of proteins longitudinally in DMD patients, proving evidence that in-depth information over the pathophysiologic pathways can be monitored over time using non-invasive methodologies. We were able to identify significant changes in protein concentrations over time in longitudinal samples despite the low patient numbers thanks to the relatively large effect sizes observed in the DMD population and low intra-assay variation of SOMAscan data. We believe that this study provides a list of candidate biomarkers to be studied in more detail prospectively in natural history studies, which may help investigators design clinical trials and evaluate patients’ progression and response to therapy in dose findings studies.

### Online supplementary material

Additional Supporting Information may be found online in the supporting information tab for this article.

- **Table S1.** Characteristics of the DMD patients involved in the cross-sectional proteomic study.
- **Table S2.** Characteristics of the DMD patients involved in the longitudinal proteomic study.
- **Table S3.** List of the 285 proteins differentially represented between DMD patients and healthy controls.
- **Table S4.** List of the significant pathways identified by over-representation analysis in the cross-sectional proteomic comparison between DMD and healthy individuals and in the longitudinal study.
- **Table S5.** List of proteins affected over time in the longitudinal proteomic comparison. Estimated changes per year and adjusted P-values are provided.
- **Table S6.** List of proteins identified to be associated with an increased risk of wheelchair dependency. The association was not significant after multiple testing correction.

### Conflict of interest

None.

### Acknowledgements

We are grateful to Biomarin for providing us with serum samples of paediatric healthy controls. We acknowledge the support by the European Community’s Seventh Framework Programme (FP7/2007–2013) under grant agreements 305121 ‘Integrated European Project on Omics Research of Rare Neuromuscular and Neurodegenerative Diseases (NEUROMICS)’, 305444 ‘RD-Connect’ and 601573 SCOPE-DMD. P.S. and A.A.R. would like to acknowledge the Association Française contre les Myopathies (grant n. 17724), ZonMw (grant n. 113302001) and Parent Project Muscular Dystrophy (PPMD) for supporting this work. F.M. is supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London. The authors certify that they comply with the ethical guidelines for authorship and publishing of the Journal of Cachexia, Sarcopenia, and Muscle.\textsuperscript{47}
References


37. Krause C, Kloen P, ten Dijke P. Elevated transforming growth factor β and mitogen-activated protein kinase pathways mediate...


