EXAMENSARBETE INOM BIOTEKNIK,
AVANCERAD NIVÅ, 30 HP
STOCKHOLM, *SVERIGE 2017*

Exploring genotype to phenotype correlations in Duchenne muscular dystrophy

CAMILLA JOHANSSON

KTH
SKOLAN FÖR BIOTEKNOLOGI

Exploring genotype to phenotype correlations in Duchenne muscular dystrophy

Master's Thesis

Camilla Johansson

Master's thesis project in Medical Biotechnology
KTH ROYAL INSTITUTE OF TECHNOLOGY
May 26, 2017, Stockholm, Sweden

Supervisor: Cristina Al-Khalili Szigyarto
Examiner: Peter Savolainen

Abstract

Duchenne muscular dystrophy, DMD, is a rare, X-linked recessively inherited genetic disorder which cause progressive muscle degradation in young boys. The disorder is caused by a number of different mutations in the dystrophin gene which disturbs the reading frame and prevents production of a functional dystrophin. Patients are known to lose the ability to walk before the age of 12, with some milder cases of patients who remain ambulant until their late teenage. In this project, a molecular approach was taken to explain the difference in disease severity on a genetic level, as well as to investigate correlations between mutation, dystrophin profiles in blood and disease severity. Patient data and serum samples from DMD patients were obtained from three different cohorts within the BIO-NMD consortium (www.bio-nmd.eu). Dystrophin levels in serum were measured using an antibody suspension bead array targeting different parts of dystrophin. Isoform distribution was detected with Western Blot. Different statistical strategies were employed to screen for correlations between genotype, protein profiles and clinical phenotype. The analysis revealed no general correlations between genotype and disease severity for explaining the phenotype in all patients, but for a few specific mutations, mechanisms such as exon skipping events were proposed to contribute to a milder disease progression. This hypothesis provided a genetic explanation for the clinical phenotypes observed in certain patients, but could not be confirmed when analyzing their respective dystrophin levels in serum. Despite that, elevated levels of C-terminal dystrophin isoforms in serum were proposed as a potential biomarker for DMD, as it was observed that patients diagnosed with DMD had in general higher levels of some dystrophin isoforms in serum compared to control samples. Also, an isolated case of an in-frame deletion of exons 5-44 suggested that internally deleted dystrophin molecules could be detected in serum from patients with in-frame deletions. For this isolated case, both the age of loss of ambulation and the dystrophin protein profile in serum could be explained by the patient's genotype. This case suggests that serum samples might provide an alternative to muscle biopsies when monitoring the effect of new gene-based therapies for treating DMD.

Sammanfattning

Duchenne muskeldystrofi, DMD, är en ovanlig, X-kromosomassocierad, recessivt ärvd genetisk sjukdom som drabbar unga pojkar och leder till progressiv nedbrytning av skelettmuskler. Sjukdomen orsakas av mutationer i genen som kodar för dystrofin, vilka förskjuter läsramen för genen och därmed förhindrar produktion av fungerande dystrofin. Patienter förlorar oftast helt förmågan att gå innan 12 års ålder. Det förekommer dock mildare fall där patientens förmåga att gå vidhålls till senare tonåren. I detta projekt har en molekylär approach antagits för att förklara denna skillnad i sjukdomsseveritet genom att undersöka korrelationer mellan mutation, dystrofin-nivåer i serum samt kliniska parametrar som t.ex. åldern då patienten förlorade förmågan att gå. Patientdata samt serum från DMD patienter tillhandahölls från tre olika kohorter inom BIO-NMD konsortiet (www.bio-nmd.eu). Dystrofin-nivåer i serum mättes med hjälp av en lösnings-baserad matris av antikroppar kopplade till magnetiska kulor (eng. antibody suspension bead array), där matrisen var utformad för att detektera olika delar av dystrofin. Distributionen av olika dystrofin-isoformer i serum detekterades med Western Blot. Olika statistiska strategier utnyttjades för att söka efter samband mellan genotyp, proteinprofiler i serum samt klinisk fenotyp. Analyserna kunde inte påvisa några globala samband mellan genotyp och sjukdomsseveritet för all patienter. Däremot föreslogs flera genetiska mekanismer, däribland exon skipping, kunna bidra till ett mildare sjukdomsförlopp för patienter med vissa specifika mutationer. Dessa teorier kunde dock inte bekräftas i serum, då mängden dystrofin inte skiljde sig signifikant mellan mild och svår fenotyp. Förhöjda nivåer av C-terminala dystrofin-isoformer kunde dock mätas i serum från patienter med DMD jämfört med kontrollpatienter, och därmed föreslogs det att vissa dystrofin-isoformer kunde fungera som biomarkörer för DMD. Dessutom observerades ett isolerat fall av en patient med en läsrams-bevarande deletion av exon 5-44, där intärnt deleterat dystrofin kunde detekteras i serum. För detta isolerade fall kunde både sjukdomsseveriteten samt dystrophinprofilen i serum direkt förklaras av patientens genotyp. Detta fall påvisar därmed möjligheten att använda serum som en alternativ analysmetod till de muskelbiopsier som idag används bland annat för att detektera effektiviteten av nya, gen-baserade behandlingsmetoder mot DMD.

Table of Contents

Abstract i

Sammanfattning ii

1. Introduction 1

1.1. The dystrophin protein and the dystroglycan complex 2

1.2. Dystrophin isoform expression in different tissues 2

1.3. Current methods for diagnosis and future aspects of disease 3

2. Materials and Methods 4

2.1. Collection of samples and patient data 4

2.2. Mapping of mutations to exons, protein domains and isoforms. 4

2.3. Screening for correlations between genotype and clinical phenotype 5

2.4. Analysis of antibody suspension bead array data from an earlier biomarker discovery trial using the UNEW cohort 5

2.5. Multiplexed antibody bead array analysis of serum samples 6

2.6. Western Blot analysis of selected serum samples 7

2.7. Prediction of dystrophin isoform expression in patients 8

3. Results 8

3.1. Mutation map 9

3.2. Screening for correlations between genotype and clinical phenotype 11

3.3. Analysis of antibody suspension bead array data of cohort UNEW from an earlier biomarker discovery study 12

3.4. Antibody suspension bead array analysis of one longitudinal cohort 14

3.5. Western blot analysis of selected serum samples 15

4. Discussion 16

5. Future perspectives 20

6. Acknowledgements 20

7. Reference 21

1. Introduction

Rare, genetic disorders, although individually rare, are collectively affecting millions of individuals worldwide. The disorders usually concern mutations within a single gene, but the consequences of those mutations can prove fatal for the affected individual [1]. Development of new treatments and diagnostic tools for these patients may require understanding of how the mutated gene affects the expression and function of its protein product on a molecular level, and to correlate this to the symptoms displayed. However, obtaining biopsy samples from the affected tissues for protein expression analysis is not always feasible, especially if young children are involved. Therefore, the possibility of using alternative, less invasive samples such as blood or urine have been considered [2]. Blood has the advantage that it circulates the entire body, and tissue-specific proteins might leak into this circulation form damaged tissues.

Duchenne muscular dystrophy (DMD) is a rare X-linked, recessively inherited genetic disorder [3] which affects about 1:3500 young boys every year [4]. The disorder is caused by mutations in the gene encoding the dystrophin protein, which is expressed in a number of different cells, including skeletal- and cardiac muscle cells [5]. In skeletal muscle cells, dystrophin is known to acts as a flexible bridge between the sarcolemma (plasma membrane) and basal lamina of the extracellular matrix, protecting the muscle fibres from breaking due to long-term contraction-induced damage [5]. In patients with DMD, this protection is absent or severely reduced, as their cells lack dystrophin production or exhibit mosaic pattern expression [6] with reduced dystrophin function. The patients suffers from progressive dystrophy of their skeletal muscles, usually losing their ability to walk before the age of twelve. The disorder eventually leads to respiratory failure or heart failure and many patients die in their early 20s [7].

A milder form of muscular dystrophy, Becker muscular dystrophy (BMD), is also caused by mutations in the dystrophin gene. BMD patients have an overall less severe disease progression than DMD patients, with production of partially functional dystrophin molecules in their cells. Many remain asymptomatic until late in life, and ambulant for another 15-20 years after the first symptoms appears [5, 7]. The difference in disease severity between DMD and BMD patients can partly be explained by the so called reading-frame rule [7, 8], where DMD is caused by mutations (mostly deletions or duplications) which disrupts the reading-frame and cause premature translation termination. BMD is usually caused by mutations which does not affect the reading frame. However, there are reported cases [7] which deviate from the reading-frame rule. Among these are certain out of frame deletions and duplications which, while causing DMD in most patients, give rise to the milder BMD in other. A possible explanation for this is the existence of exon skipping events in these BMD patients, where alternative splicing of pre-mRNA causes the mutated exons to get spliced out which restores the reading frame [9]. Other explanations are the existence of an internal ribosomal binding site both in exon 5 and exon 8, which may rescue the dystrophin production for patients carrying mutations prior to those, and the functional importance of the protein domain or subdomain which has been mutated [5, 10]. However, the severity of disease within the two disorders has been much more complex to assess on a molecular level [5, 11].

An earlier study conducted at KTH aimed at evaluating potential biomarkers for DMD in serum and plasma samples from DMD patients in the BIO-NMD consortium (www.bio-nmd.eu) [2]. The study revealed that dystrophin could be detected in blood samples from certain DMD patients (unpublished data). In this project, serum samples from that study are further analysed as an attempt to investigate the effect of different mutations in the dystrophin gene upon the disease severity and progression in DMD patients. The working hypothesis is that the severity of DMD is caused not only by the type of mutation

and its location on the dystrophin gene but also by its effect on the protein product. It is further suggested that the dystrophin molecules detected in blood were truncated dystrophin leaking into the circulation from damaged muscle cells or other tissues. Hence, there might be a relationship between the type of dystrophin molecules present in blood and specific aspects of the clinical phenotype, such age of complete loss of ambulation or rate of loss of respiratory function.

1.1. The dystrophin protein and the dystroglycan complex

Dystrophin is a 427 kDa protein. It consists of four major functional domains [11]; the actin-binding N-terminal domain (encoded by exon 1-8) which interacts with γ-actin filaments, the central rod domain (encoded by exon 8-61) which acts as a bridge between the actin filament and the sarcolemma, the cysteine-rich domain (exon 62-69) interacting with the β-dystroglycan at the plasma membrane (sarcolemma), and the carboxyl-terminal (C-terminal) domain (exon 69-79). The actin-binding domain consists of two calponin-homology (CH) domains, CH1 and CH2 [11]. Studies have shown that the amino-terminal (N-terminal) domain is not essential for dystrophin function, as mice models lacking the actin-binding domain exhibit no or very mild symptoms. This could be explained by the evidence of a second actin-binding domain in the middle third of the dystrophin rod domain [12]. The central rod domain consists of 24 spectrin-like repeats separated by four hinges. Apart from containing the second actin-binding domain, it also interacts with intermediate filaments, microtubules, the muscular isoform of nitric oxide synthase (nNOS) and membrane phospholipids. The cysteine-rich domain is composed of a WW domain, two EF handles and a ZZ domain [11], and together with the C-terminal domain it interacts with a number of different proteins including β-dystroglycan, syntrophin and dystrobrevin [11], which together makes up the dystrophin-glycoprotein complex [12]. A graphical illustration of the domains in dystrophin can be seen in Figure 1D.

1.2. Dystrophin isoform expression in different tissues

The dystrophin gene is the largest gene in the human genome [5]. It has three unique first exons with their own promoters as well as 78 common exons, allowing three slightly different, full-length isoforms to be expressed. The three promoters are tissue dependent. The brain promoter drives expression primarily in cortical neurons and the hippocampus of the brain, while the Purkinje promoter is expressed at high concentration in the cerebellar Purkinje cells and to some extent in skeletal muscle cells. The skeletal muscle cells have their own promoter, the muscle promoter, which drives expression in skeletal muscle cells and cardiomyocytes [5].

There are also protein evidence of at least four internal promoter regions, which allows the production of shorter dystrophin isoforms lacking the actin-binding domain. These shorter proteins use unique first exons which are spliced into exon 30, 45, 56 and 63. The resulting proteins have molecular weights of 260 kDa (named Dp260), 140 kDa (Dp140), 116 kDa (Dp116), and 71 kDa (Dp71), respectively. Their expression is tissue specific and none of the isoforms have been detected in skeletal muscle cells [5]. There have, however, been studies linking the lack of Dp140 expression in some DMD patients to cognitive impairment [13]. Although the literature usually only describe seven protein coding isoforms, the Ensembl.org automatic annotation software predicts the presence of at least 20 protein coding splice variants of the dystrophin gene [14].

1.3. Current methods for diagnosis and future aspects of disease

Today, diagnosis of DMD in young boys is performed through a combination of physical testing to detect weakness of proximal muscles, genetic testing to detect mutations in the dystrophin gene, immunohistochemical staining of dystrophin in muscle biopsy samples, and measurement of creatine kinase (CK) levels in blood samples [2, 15]. Following diagnosis, which usually occur at the age of five, there are a few options for medical treatments available, where the historically most common treatment has been administration of corticosteroids [15]. These drugs leads to an increased muscle strength, prolonging the time when the patient remains ambulant by 2-5 years. However, corticosteroid treatment does not restore dystrophin function and the treatment is associated with several side effects, including excessive weight gain which may lead to vertebral fractures [15]. Today, a number of new treatments are under clinical trials, involving exon skipping therapies [16], gene replacement therapies [17] and premature stop-codon read-through therapy [15]. The idea behind exon skipping therapies is to induce the ribosome to skip translation of the mutated exons, restoring the production of a functional, internally deleted dystrophin molecule. This converts the disorder from the severe DMD phenotype to a phenotype resembling the milder BMD disorder [16]. As of 2016, the first exon skipping therapy for treatment of DMD has been granted an accelerated approval by the U.S Food and Drug Administration [18].

Current methods for monitoring the dystrophin production in DMD patients pre- and post-exon skipping therapy involves analysis of muscle biopsy samples [16], a procedure which is painful for the children that are being treated. If this project could prove a relationship between patient genotype and dystrophin levels in blood samples, it is proposed that analysis of blood samples could provide a less invasive, systemic alternative to muscle biopsy sampling when monitoring the effect of exon skipping therapies.

In conclusion, the working hypothesis is that the severity of DMD is caused not only by the type of mutation and its location on the dystrophin gene but also by its effect on the protein products. If the disease progression within DMD is the result of the expression or lack of expression of not only the full-length skeletal muscle isoform, but a combination of multiple dystrophin isoforms, then this difference could possibly be detected when measuring dystrophin levels in blood samples. Patient data and serum samples from three different clinical sites within the BIO-NMD consortium was used to investigate possible correlations between genetic mutation, clinical parameters and dystrophin levels in serum from patients diagnosed with DMD.

2. Materials and Methods

2.1. Collection of samples and patient data

DMD patient data from three different sites, University College London (UCL), Leiden University Medical Center (LUMC), and Newcastle University (UNEW), was provided through the BIO-NMD consortium. The cohorts contained 44, 90 and 66 DMD patients, respectively. From the patient data, the following parameters were extracted: center, patient ID, date of birth, date of last clinical visit (or age of last clinical visit), mutation annotation, mutation type, clinical diagnosis, wheelchair use, age (or date) of complete loss of ambulation (when applicable), sample codes for each sample, date of sample collection, sample type, FVC % of predicted, NSAA performed (yes or no), latest NSAA score, corticosteroid use/type, heart medication. Patient information was anonymized.

All patients had approved the use of biological samples and patient data for research within DMD. Samples were collected according to standardized procedures. The control samples provided from each cohort were either female carriers of DMD, female carriers of BMD, and patients with disorders other than DMD or BMD. Thus, it is important to note that none of the control samples from these cohorts were retrieved from healthy individuals. There were, however, six external, longitudinal serum samples available at KTH, derived from two different, healthy individuals. These individuals lacked patient data, but had approved the use of biological samples for research purposes.

2.2. Mapping of mutations to exons, protein domains and isoforms.

The provided information about dystrophin mutations in the patient data files were recorded according to standardized HGVS-nomenclature [19] for DNA- and protein sequence variants, and the type of genetic test used to assess the genotype was not stated. Since no DNA sequencing files (e.g. FASTA-files) were available for the mutations, each mutation was visually mapped to exon positions in the dystrophin gene by manually sketching the DNA sequence and mutations in a vector based drawing program. The transcript ENST00000357033.8 from Ensembl.org [14] was used as reference for exon sequences and number. As only the last 31 nucleotides in exon 1 of transcript ENST00000357033.8 are translated, nucleotide numbering was defined so that the first nucleotide in exon 2 would have number 32. Each exon start position was then calculated by combining the start position of the previous exon with the length of the previous exon. This method generated a similar nucleotide numbering as used by the Leiden DMD Mutation Database [7]. A vector based drawing program, Inkscape, was used to make a made to scale sketch over the cDNA of the dystrophin gene and to map all mutations of the three cohorts to the gene. Transcripts for predicted protein coding splice variants of the dystrophin gene (table 1) [14] were mapped to the gene by using transcript ENST00000357033.8 as reference sequence. Protein domains and their position on the primary structure of the dystrophin protein were obtained from the NCBI Reference Sequence NP_000100.2 for the dystrophin Dp427c isoform [20].

The amino acid sequences for the antigens targeted by the antibodies HPA002725 and HPA023885 (Atlas Antibodies) were mapped to the protein sequence of transcript ENST00000357033.8 from Ensembl.org [14] and exon region of dystrophin using pairwise sequence alignment (Needleman-Wunsch alignment algorithm, substitution scoring matrix =BLOSUM62, gap open penalty = 10, gap extension penalty = 0.5) with the EMBOSS NEDLE tool from EMBL-EBI [21]. Four monoclonal anti-dystrophin mouse-IgG antibodies, MANDYS129 [22, 23], MANDYS131 [22, 23], MANEX7B [24] and MANEX58A [25], were mapped to the dystrophin gene using information about the exons and/or amino acid sequences their

antigens corresponded provided in the data sheet of each antibody (Glenn Morris, Wolfson Centre for Inherited Neuromuscular Disease, U.K.).

Table 1. Predicted protein coding splice variants of dystrophin enlisted in Ensembl.org.

Name	Transcript ID	Bp
DMD-001	ENST00000357033.8	13956
DMD-006	ENST00000378677.6	13932
DMD-203	ENST00000378707.7	7410
DMD-204	ENST00000541735.5	7080
DMD-201	ENST00000343523.6	5623
DMD-013	ENST00000378723.7	4645
DMD-015	ENST00000378702.8	2617
DMD-014	ENST00000361471.8	2224
DMD-206	ENST00000620040.4	13746
DMD-205	ENST00000619831.4	13734
DMD-202	ENST00000359836.5	7339
DMD-017	ENST00000358062.6	6773
DMD-016	ENST00000474231.5	5209
DMD-002	ENST00000288447.8	3856
DMD-019	ENST00000378680.6	1858
DMD-020	ENST00000378705.3	714
DMD-024	ENST00000493412.1	616
DMD-009	ENST00000447523.1	375
DMD-004	ENST00000420596.5	303
DMD-005	ENST00000448370.5	123

2.3. Screening for correlations between genotype and clinical phenotype

All plots were constructed in R [26] using the packages ggplot2 [27] and ggsignif [28]. All mutations in the cohort were assumed to follow the reading-frame rule, if not otherwise stated. This meant that each mutation would cause a frameshift of the reading frame, resulting in a premature translation termination within 30 nucleotides from the mutation. Hence, the exon in which a mutation start (e.g. first deleted exon) was chosen as an estimation for where premature translation termination would occur. This simplification was assumed to hold for all deletions, small deletions and nonsense mutations, but not for duplications, missense mutations and splice site mutations.

2.4. Analysis of antibody suspension bead array data from an earlier biomarker discovery trial using the UNEW cohort

Two different antibodies were used. The antibodies were polyclonal rabbit anti-dystrophin IgG developed by the Human Protein Atlas (www.proteinatlas.com), HPA002725 (Atlas Antibodies) and HPA023885 (Atlas Antibodies). Data from an early stage of a biomarker discovery study [2] was made available for this project.

The net mean fluorescence intensity (Net MFI) obtained in the study for both antibodies were combined with the available patient data from that cohort. The Net MFI for both antibodies were plotted against different clinical parameters, mutation type and mutation location. All plots were made in R [26] suing the packages ggplot2 [27], ggsignif [28] and hmisc [29].

2.5. Multiplexed antibody bead array analysis of serum samples

64 serum samples from a longitudinal cohort of DMD patients, provided by LUMC in December 2014, together with six longitudinal control serum samples from two different patients, were analysed to determine if three different parts of the dystrophin protein could be detected in blood from DMD patients with known genotype, age and clinical phenotype.

Sample labelling: Samples were thawed on ice. For each sample, 22 µl PBS (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 1.46 mM KH_2PO_4, pH 7.4) was mixed with 3 µl serum sample and 5 µl 1:2000 Biotin solution (2 mg EZ-link NHS-PEG_4-Biotin, 200 µl DMSO, 330 µl PBS), sealed, vortexed and incubated for 2 h, 4°C, 600 rpm. The samples were vortexed every 30 min. Reaction was stopped by adding 12.5 µl 0.5 M Tris-HCl (pH 8.0), followed by 20 min incubation, 650 rpm, R.T. Samples were stored in -20°C for 2 days.

Antibodies: In total, six antibodies targeting three regions of dystrophin were used (see Figure 1E). Two of the antibodies were polyclonal anti-dystrophin rabbit-IgG, HPA002725 (Atlas Antibodies) and HPA023885 (Atlas Antibodies) and four were monoclonal anti-dystrophin mouse-IgG, MANDYS129 [22, 23], MANDYS131 [22, 23], MANEX7B [24] and MANEX58A [25]. The monoclonal anti-dystrophin mouse IgG antibodies were donated from Wolfson Centre for Inherited Neuromuscular Disease (CIND), United Kingdom. A polyclonal anti-carbonic anhydrase 3 rabbit IgG, 0121, was used as a positive control. The antibodies HPA002725, HPA023885, MANDYS129, MANDYS131, MANEX7B, MANEX58A and 0121 were separately diluted in 0.1 M MES (pH 4.5) to get a total of 1.75 µg and a volume of 100 µl. The volumes used for each antibody is shown in table 2. As negative controls, a mix of serum from each patient in the cohort, and bovine serum albumin, were diluted as shown in table 2.

Table 2. Dilution of antibodies and negative control proteins in MES buffer to a volume of 100 µl/reagent before coupling to MagPlex microspheres.

Antibody	Antigen	Dilution when added to beads
HPA002725	Dystrophin, exon 7-10	43.5:100
MANEX7B	Dystrophin, exon 7/8	1.5:2000
MANDYS129	Dystrophin, exon 38/39	1.8:2000
MANDYS131	Dystrophin, exon 40/41	1.5:2000
HPA023885	Dystrophin, exon 57-59	3:100
MANEX58A	Dystrophin, exon 58	1.5:2000
0121	Carbonic anhydrase 3	29:100
Negative control protein		**Dilution or concentration when added to beads**
Serum mix (mix of serum from all patients in cohort)		1:10
Bovine serum albumin		0.175 mg/ml

Coupling of antibodies to beads: 10 different MagPlex® Microsphere mother bead tubes with known bead IDs where selected and vortexed. 40 µl of each bead ID was transferred to a separate well in a 96 well Grainger assay plate with U-bottom. The plate was placed on a magnet separator (LIFESEP® 96F Magnetic Separation Unit, Dexter Magnetics, 2501008-1) and beads washed by removing supernatant on magnet and adding 80 µl activation buffer (AB, 0.1M NaH_2PO_4, pH 6.2) OFF magnet, sealing plate and dissolving beads through vortexing. The supernatant was again removed ON magnet and 50 µl AB added OFF magnet. 50 µl of freshly prepared 52 mM N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (Sigma-Aldrich, E7750-1G) 46 mM N-Hydroxysulfosuccinimide (Thermo Scientific, 24510) solution in AB was added to each well directly following preparation, and plate was vortexed and incubated

in dark for 20 min on plate shaker, R.T at 600 rpm. Following incubation, plate was washed in 2 cycles x100 µl 0.1 M MES buffer, pH 4.5, ON magnet. Each wash cycle was performed by letting the beads settle on the magnet for approximately 30 s., removing the supernatant ON magnet using a pipette and re-suspending the beads in the wash volume OFF magnet. Directly after the two wash cycles, the nine 100 µl antibody- and control dilutions were added to the beads, one antibody or control dilution per bead ID. The tenth bead ID was dissolved in 100 µl 0.1 M MES buffer, hence not coupled to any protein. Plate was sealed and incubated in dark for 2 hrs at R.T, 650 rpm. Following incubation, the plate was washed in PBS-T ON magnet (PBS 0.05 % v/v Tween 20) for 2 cycles a 100 µl. 50 µl PBS 5 % w/v Bovine Serum Albumin (Albumin fraction V from bovine serum, Merck KGaA, 1.12018.0100) 0.05 % v/v Tween 20 was added to each well OFF magnet. Plate was sealed, vortexed and incubated in dark for 2 hrs at R.T, 650 rpm.

Coupling efficiency test: The coupling efficiency of antibodies/proteins to beads was tested by mixing 10 µl of each bead ID into a bead stock, vortex and then transfer 5 µl bead stock into three separate wells on a U-bottomed 96 well assay plate. 50 µl 1:2000 dilution of secondary detection antibody (goat anti-rabbit IgG –R-Phycoerythrin, Sigma-Aldrich, P9795-.25ml) in PBS-T was added to each well. Plate was vortexed and incubated for 20 min in dark, at 650 rpm, R.T, followed by three wash cycles a 100 µl PBS-T ON magnet. After the last wash, the beads were dissolved in 100 µl PBS-T and analyzed using the Luminex® 100/200™ System with the xPONENT® software.

Sample assay run: Biotinylated serum samples were collected from -20°C, thawed on ice for about 30 min and centrifuged, 2000 rpm for 30 s. For each sample in the cohort, 1 µl biotinylated sample was mixed with 50 µl PBS-T 10 % v/v bovine IgG (lyophilized bovine IgG from serum, Sigma Chemical Co., I-5506) 1:1000 ProClin™ 300 (Sigma-Aldrich, 48912-U) in a PCR plate. The diluted samples were heat treated in a PCR machine (56°C for 30 min, 23°C for 10 min). A bead stock was prepared by mixing 36 µl of each protein-coupled bead ID. From this bead stock, 5 µl was added to each sample after heat treatment, and plate was vortexed before incubating O/N in dark at 4°C and 650 rpm. After incubation, the plate was washed ON magnet for three cycles a 50 µl PBS-T/well, followed by re-dissolving the beads in 50 µl 1:750 dilution of streptavidin R-phycoerythrin conjugate (Life technologies, S866) in PBS-T. Plate was sealed, vortexed and incubated in dark for 20 min at 650 rpm, R.T, followed by 3 wash cycles a 50 µl PBS-T ON magnet. The beads were re-dissolved in 60 µl PBS-T and analyzed in the Luminex® 100/200™ System using the xPONENT® software. The protocol was set to measure 75 µl/well for a minimum bead count of 100 beads per ID.

2.6. Western Blot analysis of selected serum samples

6 µl of each serum sample was mixed with 12 µl of a 5xRed and glycerol mixture (2 parts 5xRed and 1 part glycerol) into separate wells on a PCR plate. 2x 6 µl milliQ and 12 µl of the 5xRed and glycerol mixture were also mixed in separate wells. The plate was vortexed and heat-treated for 6 min in 95°C. The heat treated samples were loaded onto NuPAGE™ 3-8% Tris-Acetate Protein Gels, 1.0 mm, 12-well (ThermoFisher Scientific, EA03752BOX) in replicates, where 9 µl of each sample was loaded onto each gel. 7 µl PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa (Thermo Fisher Scientific, 26619) was loaded onto two separate wells on each gel. The gels were run in an Xcell SureLock™ MiniCell Electrophoresis system (NOVEX® Life technologies, EI0001) using 1xMES running buffer (50mM MES, 50 mM Tris-Base, 0.1 % w/w SDS, 1 mM EDTA) for 2 hrs at constant 100 V. Proteins from gels were transferred to separate Trans-Blot TurboTransfer Mini 0.2 µm PVDF membranes (Bio-Rad, 170-4158) using the Trans-Blot® Turbo™ Transfer System (Bio-Rad) with the preprogrammed protocol “High MW – 1 mini

format gel per cassette" (10 min, 1.3 A, up to 25 V). Directly after transfer, both membranes were soaked in Ponceau S solution for 10 min on shake table, followed by destaining in destilled water for about a minute. The membranes were placed in plastic folders and photographed in a ChemiDoc™ XRS+ system (Bio-Rad) using a protocol for western blot prestained markers in Image Lab software (Bio-Rad). The membranes were then further destained in destilled water for a minute and placed in blocking buffer (50 mg/ml dry milk from Semper, 10 mM Tris, 0.15 M NaCl, 0.5 % v/v Tween 20) over night at 4°C. Directly after transfer to membranes, the gels were incubated in Gelcode Bluestain Reagent for 1 hr at a shake table, room temperature. This was followed by destaining of the gel in destilled water over night, shake table at room temperature. The gel was photographed with ChemiDoc™ XRS+ system.

Western Blot: Membranes were incubated 1 hr in 3.0 ml blocking buffer in a 30 ml tube with primary antibody HPA002725 diluted 1:300 or HPA023885 diluted 1:250 in 3.5 ml blocking buffer in a 30 ml tube. After incubation, both tubes were filled up with Tris buffer saline 0.05 % v/v Tween 20, TBST (10 mM Tris, 0.15 M NaCl, 0.05 % v/v Tween 20) and quickly emptied (quick wash), followed by 3 washes á 5 min in TBST on a rock'n'roll table. Polyclonal swine anti-rabbit immunoglobulin-HRP (Dako, P0399) was diluted 1:2000 in 2x3.5 ml blocking buffer and added to each tube after discarding the TBST from the final wash. The tubes were incubated with this secondary antibody for 45 min, room temperature, on a rock'n'roll table. This was followed by another quick wash and 4x5 min washes in TBST as described before. The tubes were filled up completely in TBST after final wash, to prevent the membrane from drying. Each membrane was developed in Immobilion™ Western HRP substrate (Millipore, P90714) for about 30 seconds right before imaging. Images were aquired in a ChemiDoc™ XRS+ system (Bio-Rad) using a protocol for western blot prestained markers and one for chemilumineschence acquiring.

2.7. Prediction of dystrophin isoform expression in patients

The nucleotide sequences for all 79 exons of DMD transcript ENST00000357033.8 were downloaded from Ensembl.org [14] and imported into R [26] as a data frame with a column for exon number and one for nucleotide sequence. For the patient with an inframe deletion of exon 5-44, a text file was generated containing all nucleotides of the exons prior to exon 5 and after exon 44. The nucleotide sequence was translated to amino acid sequences using the ExPASy bioinformatics tool "Translate" [30] and the molecular weight was calculated using "Compute pI/Mw" in ExPASy [30].

For all predicted protein coding splice variants of dystrophin listed in Ensembl.org, the amino acid sequences were obtained and the molecular weight of all sequences were calculated using "Compute pI/Mw" [30].

3. Results

As the aim was to detect possible correlations between genotype and severity of disease in DMD patients, the possible effects of the genotype on its protein product was first investigated by drawing each mutation in relation to the dystrophin gene and the protein product with its respective domains. Possible splice variants were also put in relation to mutations and protein domain. The effect of the mutations on different clinical parameters were then evaluated through statistical analysis of the patient data.

In order to evaluate if the genotype was related to differential protein expression, dystrophin intensities from an earlier study of one of the three cohorts were analysed. To confirm the results from that analysis, serum samples from a second cohort were analysed using an antibody suspension bead

array. Western Blot was also performed on selected serum samples in order to determine the molecular weights of the dystrophin molecules detected.

3.1. Mutation map

A mutation map was constructed in order to get a visual representation of the mutational landscape within the cohorts and how these mutations correlate to (i) exons on the dystrophin gene, (ii) predicted protein coding splice variants of dystrophin, and (iii) the functional domains of the dystrophin protein. The mutation map (Figure 1A and B) showed that most out of frame deletions are located in the area of exons 45-55. Deletions not located in the exons 45-55 area are more often located in the N-terminal region than in the C-terminal region of dystrophin. Duplications are also more commonly located in the N-terminal than C-terminal region. Small deletions and point mutations (nonsense mutations, missense mutations and splice site mutations) appear to be evenly distributed over the gene. No mutations are located in the C-terminal untranslated region.

Most of the shorter splice variants contain the cysteine rich domain, the carboxyl-terminal (C-terminal) domain and sometimes parts of the central rod domain, but lack the actin-binding domain (Figure 1D). 64.5 % of all mutation enlisted were deletions, 10.5 % duplications, 5.5 % small deletions (< 1 exon), 10 % nonsense mutations, 1 % missense mutations, 1 % splice site mutations and 8 % other mutations (not shown), including patients where the mutation is unknown or non-specifically annotated. Among these, 28.9 % of the mutations were estimated to cause a premature stop codon which only affected the transcription of the long splice variants, DMD-001, DMD-006, DMD-206 and DMD-205, as well as the shorter N-terminal transcripts DMD-002, DMD-009, DMD-004, DMD-005 and DMD-024. 49.3 % of the mutations were estimated to cause a premature stop codon between exons 45-55, which, apart from affecting the translation of the full length splice-variants, would affect the translation of the C-terminal shorter transcripts DMD-203, DMD-204, DMD-202 and DMD-016. 3.0 % would also affect the translation of the C-terminal DMD-201 transcript, and 2.5 % affect the remaining C-terminal transcripts DMD-013, DMD-015, DMD-014 and DMD-019 (Figure 1C).

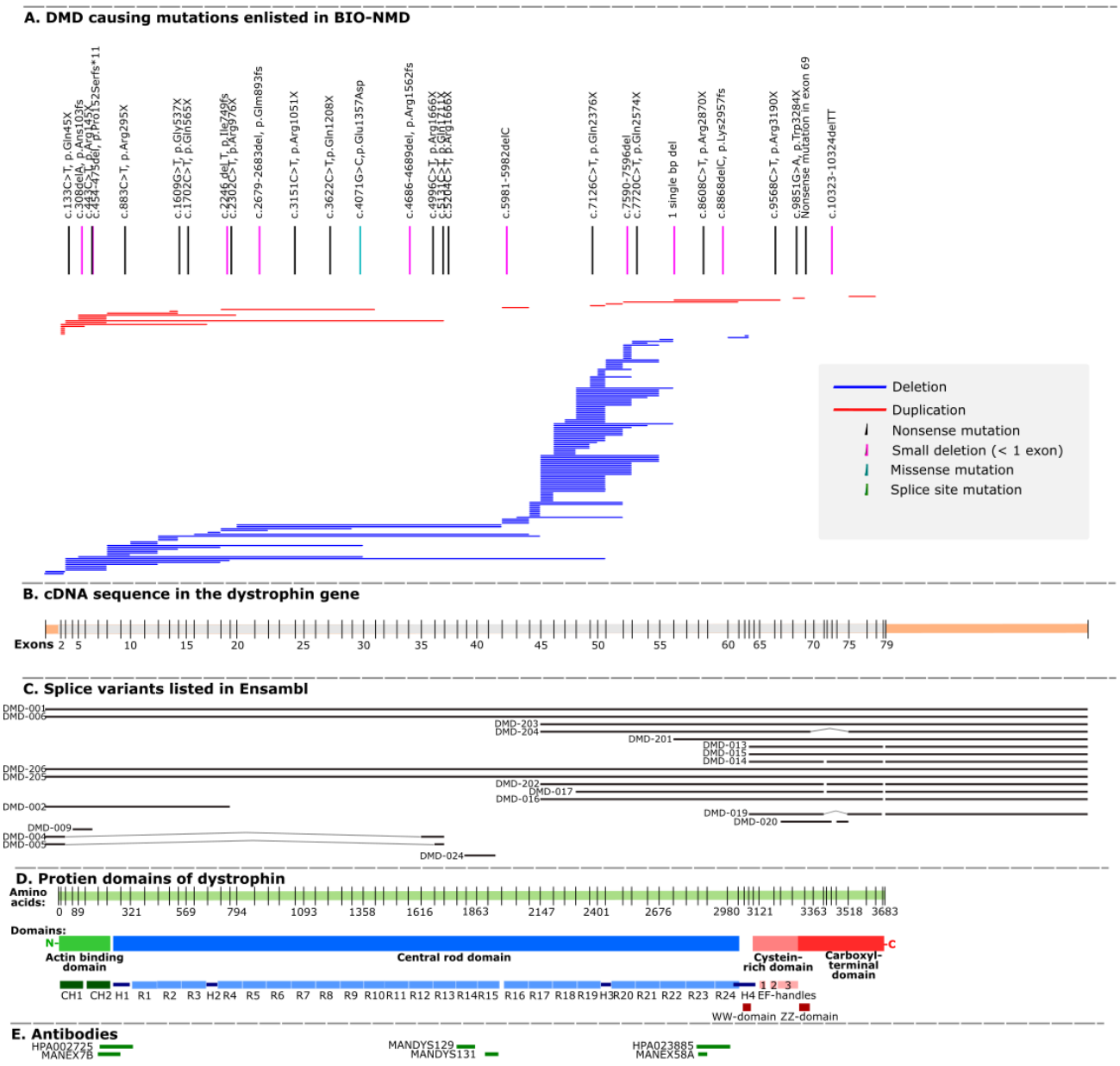

Figure 1. The dystrophin gene structure, transcripts of predicted protein coding splice variant and protein products in patients with DMD. A) Mutations within the dystrophin gene which are present in the genetic information of DMD patients in the BMD-NMD consortium. **B)** Coding DNA in the dystrophin gene, drawn using the Ensembl.org transcript ENST00000357033.8 as reference sequences. The orange part of the sequence shows non-translated regions and grey translated regions of the transcript. The vertical black lines indicates first nucleotide of each exon. **C)** Transcripts of splice variants of the dystrophin gene enlisted in Ensembl. Exon numbering in each transcript has been denoted using transcript ENST00000357033.8 (DMD-001) as reference sequence. **D)** Protein domains of dystrophin constructed from NCBI Reference Sequence NP_000100.2. CH = calponin homology domain, H = hinge domain, R = spectrin repeat domain. **E.** Recognition epitopes for anti-dystrophin IgG antibodies used in project.

3.2. Screening for correlations between genotype and clinical phenotype

In order to investigate if the mutation type and/or location for each patient could be related to any one of the clinical parameters used to monitor disease progression, the clinical parameters of interest were first selected based on amount of patients where data for this phenotype was available. The age of complete loss of ambulation was plotted against mutation type and location of the mutation in the dystrophin gene, based on the location of the first mutated exon (Appendix A). Wilcoxon rank sum tests were used as a statistical test to evaluate if two datasets in the plots were significantly different from each other. No significant correlation was shown between age of complete loss of ambulation and the location of the mutation. Patients with frameshift whole exon deletions showed no difference in age of loss of ambulation from patients with frameshift whole exon duplications, however the data suggests that patients with small frameshift deletions, covering less than 1 exon, tend to lose their ability to walk early in life, in this case close to nine years of age, while patients with nonsense mutations tend to lose their ability to walk on average a few years later. This appears to be the case independent of the location of the mutation on the gene and holds for all three cohorts. However, analysis of a larger set of patients with these mutations would be required to further confirm this observation.

The severity of disease was defined as whether or not ambulation was lost before age of 13. Hence, all patients who lost ambulation before age 13 were defined as having severe DMD (sDMD), and all patients who either had lost ambulation after age of 13 or remained ambulant while older than 13 at the time of sample collection were defined as having mild DMD (mDMD). Ambulant patients who had not yet turned 13 at the time of sample collection were defined as "unknown".

Table 3. A. Patients diagnosed with DMD who lost ambulation after the age of 13. **B.** Patients diagnosed with DMD who remained ambulant after the age of 13.

A.

Age at loss of ambulation	Mutation	Comment
13.0	deletion exon 44	
13.0	deletion exons 8-13	
13.6	duplication exon 2	
13.6	deletion exon 44	
14.0	deletion exon 45	
14.0	deletion exon 45	
15.5	duplication exons 5-7	
16.0	duplication exon 13	
16.0	missense mutation exon 29 (c.4071G>C; p.Glu 1357Asp)	
18.0	exon 51, unknown mutation	

B.

Age at last clinical visit	Mutation	Comment
13.1	duplication exon 2-16	intermediate wheelchair use
13.2	duplication exon 18-30	Ambulant
13.8	frameshift mutation exon 5 (c308delA; p.Asn103fs)	Ambulant
14.3	deletion exon 55	Ambulant
14.3	deletion exon 53	Ambulant
14.4	nonsense mutation exon 3 (c.133C>T; p.Q45X)	intermediate wheelchair use
14.5	deletion exons 44-51	Ambulant
14.7	deletion exon 44	intermediate wheelchair use
14.8	deletion exons 47-50	Ambulant
15.4	deletion exons 3-7	intermediate wheelchair use
17.7	deletion exons 42-43	intermediate wheelchair use

In an attempt to understand the tendencies of some patients to lose their ability to walk much later than others, the mutation data of all patients who lost their ability to walk after the age of 13, or who remained ambulant after that age, was extracted (**Table 3**). In general, patients who lost ambulation late in life tended to have short mutations, involving deletions or duplications of single or a few exons. Three point mutations were also present, one frameshift mutation, one nonsense mutation and one missense mutation.

3.3. Analysis of antibody suspension bead array data of cohort UNEW from an earlier biomarker discovery study

An earlier biomarker discovery study which investigated potential biomarkers for DMD in serum and plasma samples from the UNEW cohort [2] generated protein bead array data for dystrophin levels in blood. This data was used to analyse correlations between genotype and protein levels in serum, as well as correlations between protein profiles in serum and clinical phenotype.

The UNEW cohort contained serum samples from 60 DMD patients, 24 BMD patients, 13 asymptomatic female carriers of DMD and 3 asymptomatic female carriers of BMD. Of these, the age of the patient at sample collection was known for all but two patients, one DMD patient and one BMD patient. The age at complete loss of ambulation was known for 31 DMD patients. Among the remaining 29 DMD patients, 25 were marked as ambulant at the time of sample collection, one marked as intermittently using wheelchair and three marked as non-ambulant.

The net mean fluorescence intensity (Net MFI) of antibody HPA002725 and HPA023885 from an earlier antibody suspension bead array study of cohort UNEW were analysed against clinical and genetic parameters. The antibodies HPA002725 and HPA023885 were, through pairwise sequence alignment of their respective PrEST-antigen to the protein sequence of dystrophin, determined to bind to amino acids 185-333 and 2842-2992, respectively. These sequences corresponds roughly to exons 7-10 and exons 58-60. The Net MFI of each antibody was plotted against the mutation type and location (see Appendix B). The location of each mutation was defined based on which splice variants that would be affected by the mutation, so that if the prevention of expression of specific dystrophin isoforms would affect disease progression, there might be a detectable difference in the amount of dystrophin in serum for patients lacking these isoforms. However, no significant correlations ($P < 0.05$) were shown between the Net MFI of the antibodies and area which was mutated, nor the Net MFI and mutation type.

The behaviour of another clinical parameter, percent of predicted value for Forced Vital Capacity (FVC % of predicted), was investigated in relationship to Net MFI of both HPA002725 and HPA023885. FVC is a measurement of lung volume obtained through a spirometry test, and for DMD patients, this parameter has been shown to decrease in relationship to expected values in a healthy individual of the same age and sex, as muscle degradation prevents full usage of lung capacity [31]. Plotting of Net MFI of HPA023885 against the age of each patient at sample collection shows a significant, declining Spearman correlation for DMD patients, where the Net MFI appears to decline rapidly between patients aged 0 to 10 years, and less rapidly between older patients (Figure 2). A similar behaviour was observed for BMD patients within the cohort, but could not be observed in control patients consisting in asymptomatic female carriers of either DMD or BMD (Figure 2). Despite the age dependency of both FVC % of predicted and Net MFI of HPA023885 in serum, no significant Spearman correlation was shown between Net MFI of HPA023885 in serum and FVC % of predicted in DMD patients where a spirometry test had been conducted

the same day as sample collection (Appendix C). For the dystrophin N-terminal antibody, HPA002725, the Net MFI does not correlate with either age at sample collection nor FVC % of predicted (see Appendix C).

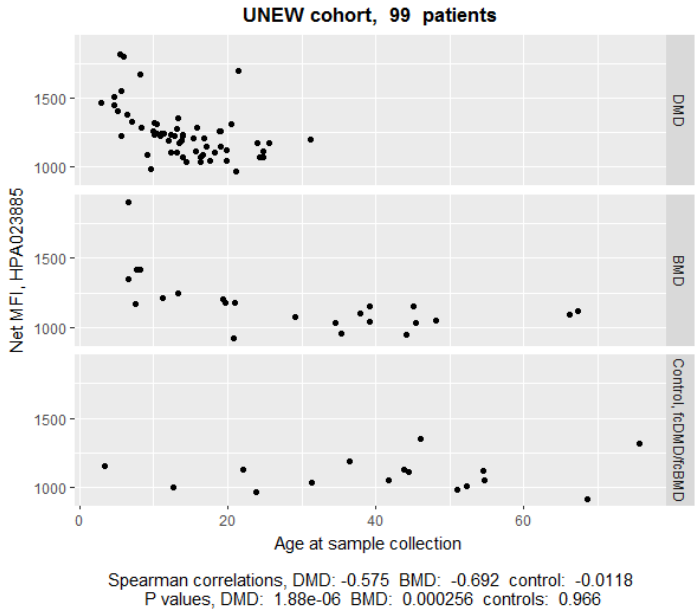

Figure 2. The Net MFI of HPA023885 plotted against patient age at sample collection. For DMD and BMD patients in the UNEW cohort, there is a significant Spearman correlation between the two parameters, where Net MFI of HPA023885 in serum declines with patient age. For control samples, consisting of serum samples from asymptomatic female carriers of either BMD or DMD, no significant Spearman correlation could be observed between Net MFI and age. The cohort consisted of serum samples from 99 patients, where 59 patients were diagnosed with DMD, 23 patients diagnosed with BMD and 16 control samples.

In order to determine if Net MFI for HPA023885 in serum was affected by the severity of DMD, the age at sample collection was plotted against Net MFI for HPA023885 for different disease severities. Figure **3**A shows the age and Net MFI of HPA023885 distribution of patients classified as either mDMD, sDMD or unknown. The unknown patients are almost all DMD patients younger than 10, who are still ambulant. There is a larger age distribution in sDMD patients than mDMD patients, but the Net MFI between the two groups is not significantly different (Figure 3B). However, the Net MFI of HPA023885 appears to be slightly higher for DMD patients older than 10 than for control patients (Figure 3B). DMD patients in this cohort were between the ages of 2.9-31.1 years at time of sample collection, with a mean age of 14.0 years and a median age of 13.4 years.

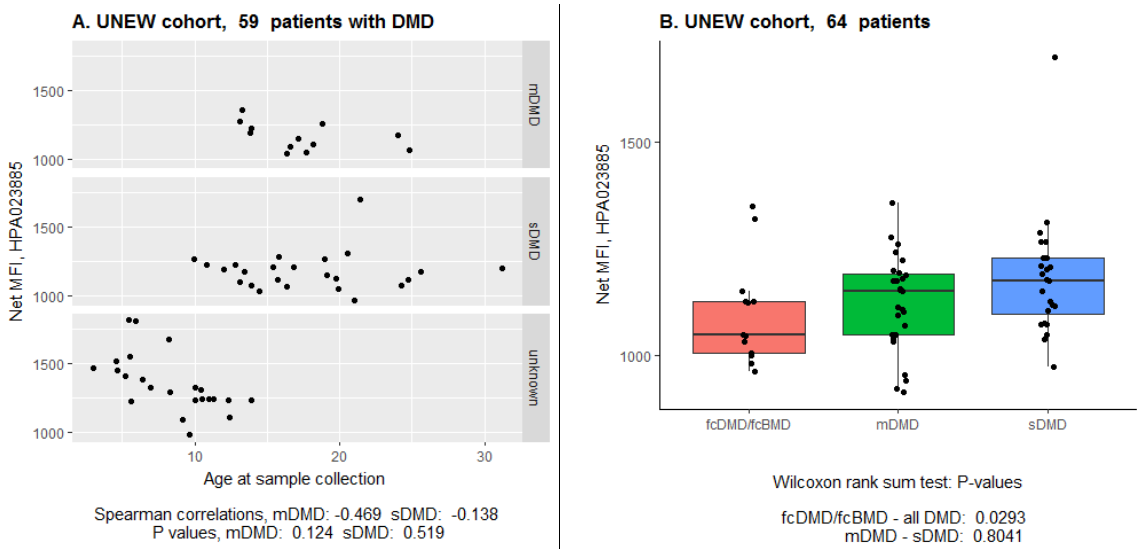

Figure 3. A. Net MFI for HPA023885 in serum plotted against age at sample collection for DMD patients with a severe (sDMD), mild (mDMD) or unknown phenotype. The disease severity is defined as whether the patients lose ambulation before or after the age of 13. **B.** The Net MFI of HPA023885 for mDMD and sDMD patients compared to control samples from female carriers of either DMD or BMD (fcDMD/fcBMD).

3.4. Antibody suspension bead array analysis of one longitudinal cohort

In order to confirm the observations from antibody suspension bead array data in the UNEW cohort, a second cohort, LUMC, was analysed using an antibody suspension bead array. Analysis of the LUMC cohort using a antibody suspension bead array with six antibodies against three different epitope areas (see **Figure 1**E) of dystrophin generated Net MFI data for 41 samples from DMD patients and 4 control samples from 2 different healthy individuals. Net MFI was defined as a median MFI value of at least 30 beads per analyte. The raw Net MFI data for all samples fulfilling this criteria can be found in Appendix D. The data revealed three samples that had Net MFI values for antibody HPA023885 which were more than 5 times higher than in other samples. These patients did not have proportionally higher Net MFI values for other analytes. Possible explanations for these signals will be considered later.

To confirm that the Net MFI of HPA023885 correlates with age, which was shown in the UNWE cohort, the Net MFI for HPA023885 was plotted against the age of each patient at sample collection (**Figure 4**). In contrast to the UNEW cohort, no correlation between age and C-terminal dystrophin products (HPA023885) could be seen. Patients between the ages of 5-10 years, which had the largest distribution of HPA023885 Net MFI in the UNEW cohort, displayed a much narrower distribution in the LUMC cohort. Also, there was no significant difference in Net MFI between DMD patients and healthy controls (see Appendix E).The age of patients at time of sample collection spanned between ages 5.3 to 17.6 years, with the mean age being 10.9 years and median 10.1 years.

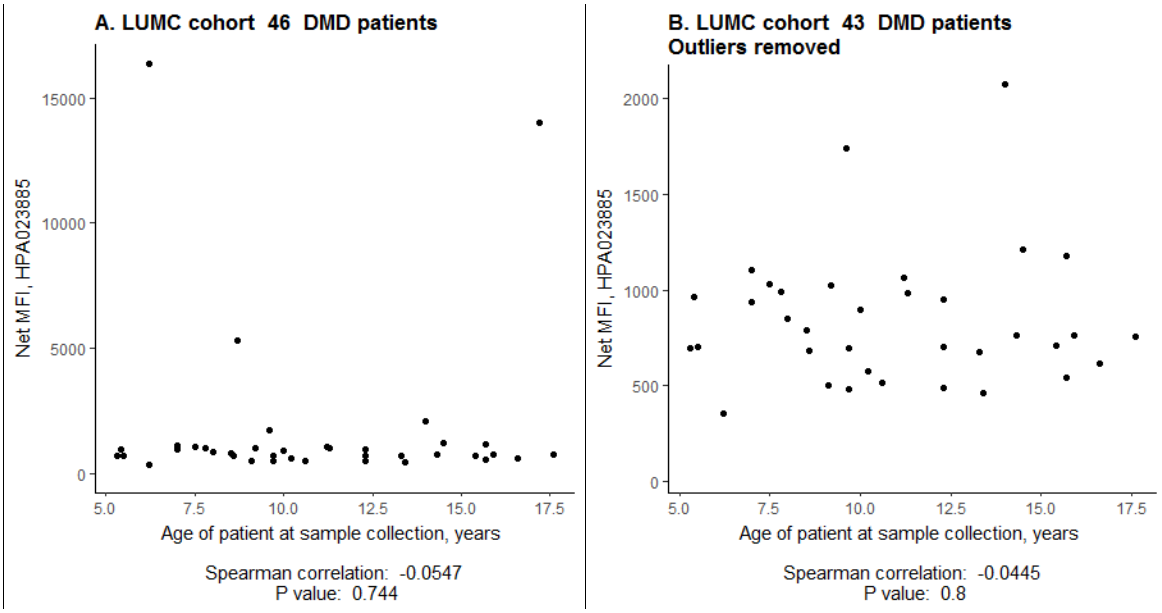

Figure 4. A. The Net MFI of HPA023885 for each DMD patient in the LUMC cohort, in relation to their age at time of sample collection. B. Three outliers with Net MFI of HPA023885 > 5000 were removed from figure 4A.

3.5. Western blot analysis of selected serum samples

The six serum samples from DMD patients in the LUMC cohort which displayed the highest Net MFI value for antibody HPA023885, and one control sample, C, were analysed using Western Blot. The primary antibodies were HPA002725, binding to the exon 8-10 area of the protein (Figure 1), and HPA023885, binding to the exon 58-60 area (Figure 1). The experiment was performed in order to investigate the size of dystrophin molecules in serum. The samples used, and the respective mutations of these patients, can be seen in Table 4. The sample numbers represents the order in which they were loaded on to the gel.

Table 4. Serum samples from DMD patients analysed by Western Blot and the respective mutations the patients have in their dystrophin gene. The numbering of the samples represent the order in which they were loaded onto the gel.

Sample number	Mutation
1	Nonsense mutation exon 27: c.3622C>T; p.Gln1208X
2	Deletion of exons 45-54
3	Deletion of exons 45-52
4	Nonsense mutation exon 36: c.5131C>T, p.Gln1711X
5	Deletion of exons 48-52
6	Deletion of exons 5-44 (inframe)
Healthy control	-

Both the membrane treated with HPA002725 as primary antibody and the membrane treated with HPA023885 as primary antibody contained a large amount of a protein or protein aggregate which stopped migrating at about 40 kDa (Figure 5A and C). This was shown by staining the membranes with Ponceau prior to blocking and antibody treatment. The membrane treated with HPA002725 showed no presence of the N-terminal region of dystrophin in any serum sample (Figure 5B), but the membrane treated with HPA023885 showed the presence of several bands at about 100 kDa, 40 kDa and 32 kDa in both DMD samples and controls (Figure 5D), as well as a very faint band at about 130 kDa. A strong, bent line could

also be seen in Figure 5D, which corresponded to the bottom line of the large protein smear visible under Ponceau staining of the same membrane (Figure 5C). Sample 6 had an additional band at about 260 kDa which is not present in any of the other DMD samples nor in the control (Figure 5D).

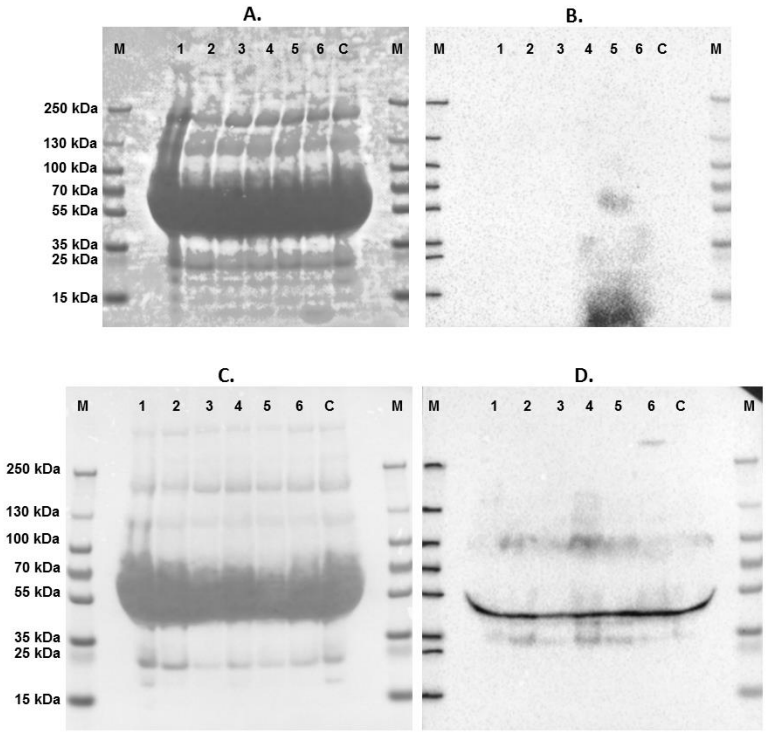

Figure 5. Analysis of dystrophin protein isoforms and fragments in serum samples. A: Ponceau staining of first membrane right after transfer and before blocking. **B:** Western Blot of first membrane incubated with HPA002725 as primary antibody. Exposure time 120 seconds. **C:** Ponceau staining of second membrane right after transfer and before blocking. **D**: Western Blot of second membrane incubated with HPA023885 as primary antibody. Exposure time was 20 seconds.

For patient 6, the possible molecular weight of a dystrophin molecule lacking exons 5-44 was calculated to be between 187 489.60 Da and 204 051.53 Da, depending on whether the entire exon 5 and 44 are deleted or if the deletion starts and/or ends within these exons. The calculated theoretical molecular weights of all protein coding splice variants can be seen in Appendix F.

4. Discussion

Duchenne muscular dystrophy has, in comparison to other rare genetic disorder, been exceptionally well studied. Despite this, the small size of the patient groups available makes it difficult to perform statistically supported analysis of the genetic and phenotypical diversity within the disorder. This turned out to be the case with the three cohorts available for this project. Most attempts to account for more than one clinical parameter at a time, or to subdivide the patients into groups based on multiple genetic and clinical parameters, reduced the amount of data points below what was possible to analyze statistically (data not included). To this comes the fact that patient data within the cohorts were collected and compiled at three different geographical sites during a time period of seven years, and the routines for data collection differed during this period. Hence, only two clinical measurements of disease severity

were concluded to be consistent enough between all three cohorts for comparative analysis. These parameters were (i) age of patient at complete loss of ambulation, and (ii) FVC % of predicted.

At the start of this project, it was hypothesized that disease progression in DMD patients could partly be explained by the type of genetic mutation the patient carried and how this mutations affected the expression of dystrophin and its isoforms in different tissues. Since the reading frame hypothesis states that about 90 % of all mutations in the dystrophin gene of DMD patients cause a frameshift [7, 8], and hence leads to a premature translation stop, it was suggested that some DMD patients might have truncated dystrophin molecules present in their skeletal muscle cells.

An analysis of the distributions of mutations within all three cohorts (Figure 1A) revealed a large distribution of frameshift causing mutations (deletions and duplications) and nonsense mutations across the gene. This was in accordance with the reading-frame hypothesis. However, there were a few missense mutations, splice site mutations as well as a single case of an in-frame deletion of exons 5-44 present in the cohorts who appear to deviate from the reading-frame hypothesis. Some of these proved interesting on a phenotypic level and will be considered in more detail later. The majority of the deletions turned out to be located between exons 45-55 (Figure 1). This area has in the literature been described as one of two deletion hotspots within the gene and known to be the most commonly mutated area in both BMD and DMD patients [5]. Overall, the distributions of mutation types, e.g. percentage of deletions and duplications, within the cohorts were in accordance with other studies [7]. The distributions were also similar between cohorts (data omitted).

Figure 1 shows that most mutations affect the central rod domain of the protein product, while leaving the actin binding domain and parts of the central rod domain intact. Such molecules could have some of the interactive functions of dystrophin preserved, and hence be involved in signaling cascades within the cells, ultimately affecting the disease progression in either a positive or negative way. Mutation that affect the actin-binding domain might instead cause the complete absence of dystrophin in skeletal muscle cells. However, there is also a contrasting hypothesis proposed in the literature [32] suggesting that the production of truncated dystrophin molecules does not by themselves provide the difference in disease severity, as the introduction of a premature translation stop codon could trigger degradation of mRNA through nonsense-mediated mRNA decay. Considering this, dystrophin production might be close to absent in some patients while partly present in others, independent of the location of a frameshift mutation. Attempts to plot the age of loss of ambulation against mutation type or location revealed no significant patterns (see Appendix A). There could be several possible explanations for this. (i) The majority of all mRNA transcripts for the long skeletal muscle isoform gets degraded in patients with frameshift mutations, and hence the location of the mutations does not mediate disease severity. Instead, the age of loss of ambulation could be affected by a combination of external factors, i.e. how active the child is and when steroid medication was initiated. (ii) The age of loss of ambulation is indeed affected by the genotype, but not as much the position or type of mutation but the occurrences of exon skipping events [9] or the use of alternative internal ribosome binding sites, IRESs, in exons 5 and 8 [5, 10]. (iii) The disease severity in DMD patients might be affected directly or indirectly by a combination of other dystrophin isoforms than the long skeletal muscle isoform. If this is the case, mutations which prevent the expression of some of the shorter dystrophin isoforms might have a larger impact on the disease severity than mutations which only affect the long isoforms. (iv) The age of loss of ambulation might be due to a combination of alternatives (i)-(iii) and the effect of all parameters combined might be too complex to assess within the scope of this project.

Table 3 shows an extract of the mutations for all patients in the three cohorts who lost ambulation after the age of 13. The table reveals that most patients with this phenotype have short mutations involving no more than a few exons. There were four patients with frameshift or nonsense mutations prior to exon 8. These patients' mild phenotype could be explained if some of their dystrophin production was rescued by utilizing the IRES in exon 8. Patients with such mutations have also been shown to sometimes exhibit the BMD phenotype [5, 10], thus it can be argued that less extensive use of this IRES could result in a less severe form of DMD. Other frameshift mutations in Table 3 known to sometimes cause BMD instead of DMD were deletions or duplications of exons 44, 45 and 51 [5]. These mutations are believed to be subject to different degrees of exon skipping in BMD patients. Hence, low occurrences of exon skipping events could explain the mild DMD phenotype as well. To further investigate this hypothesis, two patient groups were defined where one contained all patients who lost ambulation above the age of 13 (mDMD) and one with patients who lost ambulation before age 13 (sDMD). All patients who remained too young to be classified into any of the two groups were marked as "unknown" (see Figure 3A). It was assumed that if patients with the mild phenotype had an increased dystrophin production over patients with the severe phenotype, then this would be detectable in serum. However, analysis of C-terminal and N-terminal dystrophin levels in serum from DMD patients revealed no significant difference between the two severity groups (for C-terminal dystrophin, see Figure 3. For N-terminal dystrophin, see Appendix C), neither in the UNEW cohort nor in the LUMC cohort. If the disease severity was mainly mediated through alternating levels of exon skipping events or internal ribosomal binding sites, there would likely be more dystrophin present in the serum of these patients. As this was not the case, exon skipping and IRESs might not be the main cause for differences in age of loss of ambulation. Alternatively, differences in dystrophin expression are too low to be detectable in serum.

The Net MFI measured by the C-terminal antibody showed a clear age dependency within the UNEW cohort (Figure 2), while the Net MFI of the N-terminal antibody did not (see Appendix B). This suggests that the antibodies measured different proteins in the serum samples. One possible explanation could be that the N-terminal antibody measures the occasional expression of any of the full-length isoforms, either as a result of exon skipping events or through expression of truncated isoforms. In contrast, the C-terminal antibody could be measuring the presence of other, shorter isoforms in serum (see Figure 1). It was confirmed through Western Blot analysis that the C-terminal antibody measured the presence of a number of different, shorter dystrophin molecules.

The age dependency of the C-terminal antibody suggested that the levels of C-terminal dystrophin isoforms in serum is on average much higher in DMD and BMD patients aged 0-10 than in patients older than 10. Interestingly, this age dependency pattern had previously been observed for other DMD-related proteins during an earlier biomarker discovery study [2] (unpublished data). A possible explanation could be that children below 10 are growing rapidly, which increases the overall turnover of cells in the body. This increases the chances that tissue specific proteins end up in the blood stream. When the children gets older, their growth rate decreases, and hence tissue specific biomarkers becomes less abundant in serum.

The same trend was not observed in female carriers (control patients), but it is here also important to point out that the control group only contained two female carriers who were below the age of 20 at sample collection, and hence, the lack of correlation might be coincidental. Neither could the age dependency be confirmed in the LUMC cohort, as no correlation was seen between the Net MFI of the C-terminal antibody and the age of the patient (Figure 4). However, the LUMC cohort contained less samples than the UNEW cohort, with a more narrow age distribution and fewer samples for patients aged 5-10, where the largest distribution of Net MFI had been observed in the UNEW cohort. Hence, more

samples would need to be analyzed to further evaluate if the amount of C-terminal dystrophin does indeed depend on the patient's age.

In addition to the two polyclonal antibodies used when analyzing the UNEW cohort, four new, monoclonal antibodies (mAb) were also used when performing antibody suspension bead array analysis on the LUMC cohort. The array was designed to target three different regions of the dystrophin molecule with two antibodies per region (see Figure 1E). However, the Net MFI data for all four mAb suggested that the mAb-coupled beads only detected background signal (Appendix D). It was concluded that all four mAb were suspended in high horse serum supernatant, and hence the compatibility of MagPlex beads with these antibody suspensions was poor. The data from all four mAb were therefore excluded from the analysis, and all analysis were made exclusively with the N-terminal and C-terminal antibodies used in the UNEW cohort.

Two bands on the membrane, one strong at about 100 kDa and one faint at about 130 kDa could represent the dystrophin isoforms Dp116 (Mw 110 kDa) and Dp140c (Mw 129 kDa), by comparison of Appendix F and Figure 5. These were present in all six DMD samples as well as in the control sample. However, all seven samples also contained two bands at about 40 and 32 kDa, which were too small to be any of the known dystrophin isoforms. Further studies would be needed to characterize these bands, either by mass spectrometry sequencing or by the use of another anti-dystrophin antibody binding to the same region of the protein. It can neither be said which tissue or tissues these dystrophin isoforms were produced in. Although the same isoforms as in the DMD samples were also present in the control sample, and might be considered to naturally occur in the blood of all individuals, these isoforms might still have some biomarker potential for DMD. The UNEW cohort showed a significant increase in C-terminal dystrophin levels for DMD patients older than 10 compared to female carrier control samples (Figure 3).

The Western Blot also revealed a band at about 260 kDa which was only present in one patient. That patient was the only DMD patient in the cohorts who had an in-frame deletion, namely a deletion of exons 5-44. This deletion would in theory allow for production of internally deleted dystrophin in skeletal muscle cells. If so, this dystrophin molecule could leak into the blood stream when muscle cells break, and should be detectable by the C-terminal antibody. The patient had indeed the third highest Net MFI for the C-terminal antibody in the LUMC cohort, and a low Net MFI of the N-terminal antibody. The latter would be expected, as the epitope for the N-terminal antibody, encoded by exons 7-10, was deleted in this patient. Although the patient was assumed to have a dystrophin-producing skeletal muscle cells, the age of loss of ambulation was 8.1 years, which would classify the patient as a severe DMD case. An explanation to the severe disease progression, despite dystrophin production, could be that the deletion involved parts of the actin binding domain as well as the entire second binding domain located in the central rod domain [12]. Hence, the protein would lack actin binding properties which might prevent it from functioning normally. Further, the theoretical molecular weight of a dystrophin molecule lacking exons 5-44 was calculated. However, such a molecule would have a molecular weight of 187 kDa, and even if the entire exon 5 and 44 would remain intact, the molecular weight of the band on the Western Blot (205 kDa) is still a far stretch from 260 kDa. One explanation could be that the mutation causes the inclusion of sequences otherwise present in introns. Another explanation could be that the mutation is not correctly curated.

In conclusion, no general correlations between genotype and disease severity could be detected with the strategy employed. However, several explanations were proposed, including low amounts of exons skipping events, to contribute to a milder disease progression for patients with certain types of mutations. This hypothesis provides a genetic explanation for the clinical phenotypes of certain patients, but could not be confirmed when analyzing dystrophin levels in serum. However, elevated levels of C-terminal

dystrophin isoforms in serum might be a potential biomarker for DMD, and should be further analyzed. Also, an isolated case of an in-frame deletion of exons 5-44 suggested that long dystrophin molecules could be detected in serum from patients with in-frame deletions. For this isolated case, both the age of loss of ambulation and the dystrophin protein profile in serum could be explained by the patient's genotype, confirming the hypothesis that genotype to phenotype correlations can be detected in blood.

5. Future perspectives

The indicated age dependency of the concentrations of C-terminal dystrophin isoforms in serum would need to be further analyzed in other patient groups in order to determine if this behavior is characteristic for muscular dystrophy patients or not. The significant increase in concentration of C-terminal dystrophin isoforms in DMD patients compared to healthy individuals of similar age suggested that these isoforms could be considered as potential biomarkers. However, further analysis would be required to confirm this and to evaluate what information that can be obtained from such biomarkers.

Furthermore, the described case where internally truncated dystrophin was detected in serum from a patient carrying an in-frame exons 5-44 deletion opens the question of whether serum could be used to detect dystrophin production in other patients carrying deletions that preserve the reading frame. If so, detection of dystrophin isoforms in serum could be used as an alternative to muscle biopsy samples when monitoring the effect of exon skipping therapy. Since exon skipping therapies are targeted towards young children, a less invasive sampling method like serum would be desirable. Further analysis of serum from BMD and DMD patients carrying in-frame deletions is therefore proposed.

6. Acknowledgements

This project was made possible by my supervisor Christina Al-Khalili Szigyarto at the Royal Institute of Technology (Sweden) and the BIO-NMD consortium. Special acknowledgement is given to the Leiden University Medical Centre (Netherlands), Newcastle University (U.K.) and University College London (U.K.) for providing serum samples and patient data. Also, acknowledgement is given to Professor Glenn Morris at the Wolfson Centre for Inherited Neuromuscular Disease, U.K., for donating monoclonal anti-dystrophin antibodies to this project.

I would like to thank everyone at floor three, AlbaNova, for a welcoming atmosphere. A special thanks is direct to everyone in the Secretome groups at floor three, AlbaNova, for all the help and guidiance.

7. Reference

1. Boycott, K.M., et al., *Rare-disease genetics in the era of next-generation sequencing: discovery to translation.* Nat Rev Genet, 2013. **14**(10): p. 681-91.
2. Ayoglu, B., et al., *Affinity proteomics within rare diseases: a BIO-NMD study for blood biomarkers of muscular dystrophies.* EMBO Mol Med, 2014. **6**(7): p. 918-36.
3. Liang, Y., et al., *Dystrophin hydrophobic regions in the pathogenesis of Duchenne and Becker muscular dystrophies.* Bosn J Basic Med Sci, 2015. **15**(2): p. 42-9.
4. Juan-Mateu, J., et al., *DMD Mutations in 576 Dystrophinopathy Families: A Step Forward in Genotype-Phenotype Correlations.* PLoS One, 2015. **10**(8).
5. Muntoni, F.T.S.F., A., *[Review] Dystrophin and mutations: one gene, several proteins, multiple phenotypes.* Lancet Neurol, 2003. **2**: p. 731-740.
6. Helliwell, T.R., T.M. Nguyen, and G.E. Morris, *Expression of the 43 kDa dystrophin-associated glycoprotein in human neuromuscular disease.* Neuromuscul Disord, 1994. **4**(2): p. 101-13.
7. Aartsma-Rus, A., et al., *Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule.* Muscle Nerve, 2006. **34**(2): p. 135-44.
8. Monaco, A.P., et al., *An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus.* Genomics, 1988. **2**(1): p. 90-5.
9. Incitti, T., et al., *Exon skipping and duchenne muscular dystrophy therapy: selection of the most active U1 snRNA antisense able to induce dystrophin exon 51 skipping.* Mol Ther, 2010. **18**(9): p. 1675-82.
10. Wein, N., et al., *Translation from a DMD exon 5 IRES results in a functional dystrophin isoform that attenuates dystrophinopathy in humans and mice.* Nat Med, 2014. **20**(9): p. 992-1000.
11. Le Rumeur, E., *Dystrophin and the two related genetic diseases, Duchenne and Becker muscular dystrophies.* Bosn J Basic Med Sci, 2015. **15**(3): p. 14-20.
12. Ervasti, J.M., *Dystrophin, its interactions with other proteins, and implications for muscular dystrophy.* Biochim Biophys Acta, 2007. **1772**(2): p. 108-17.
13. Taylor, P.J., et al., *Dystrophin gene mutation location and the risk of cognitive impairment in Duchenne muscular dystrophy.* PLoS One, 2010. **5**(1): p. e8803.
14. Yates, A., et al., *Ensembl 2016.* Nucleic Acids Res, 2016. **44**(D1): p. D710-6.
15. Beytia Mde, L., J. Vry, and J. Kirschner, *Drug treatment of Duchenne muscular dystrophy: available evidence and perspectives.* Acta Myol, 2012. **31**(1): p. 4-8.
16. Cirak, S., et al., *Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study.* Lancet, 2011. **378**(9791): p. 595-605.
17. Mendell, J.R., L.R. Rodino-Klapac, and V. Malik, *Molecular therapeutic strategies targeting Duchenne muscular dystrophy.* J Child Neurol, 2010. **25**(9): p. 1145-8.
18. U.S. Food and Drug Administration . FDA grants accelerated approval to first drug for Duchenne muscular dystrophy. Press release. 19 September, 2016. Accessed on 22 May 2017. Available from: https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm521263.htm.
19. den Dunnen, J.T., et al., *HGVS Recommendations for the Description of Sequence Variants: 2016 Update.* Hum Mutat, 2016. **37**(6): p. 564-9.
20. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Res, 2016. **44**(D1): p. D733-45.
21. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: the European Molecular Biology Open Software Suite.* Trends Genet., 2000. **16**(6): p. 276-277.

22. Nguyen thi, M., et al., *Monoclonal antibodies against defined regions of the muscular dystrophy protein, dystrophin.* FEBS Lett, 1990. **262**(2): p. 237-40.
23. Nguyen, T.M. and G.E. Morris, *Use of epitope libraries to identify exon-specific monoclonal antibodies for characterization of altered dystrophins in muscular dystrophy.* Am J Hum Genet, 1993. **52**(6): p. 1057-66.
24. Bartlett, R.J., et al., *In vivo targeted repair of a point mutation in the canine dystrophin gene by a chimeric RNA/DNA oligonucleotide.* Nat Biotechnol, 2000. **18**(6): p. 615-22.
25. Lam, L.T., et al., *Exon-specific dystrophin antibodies for studies of Duchenne muscular dystrophy.* Translational Neuroscience, 2010. **1**(3): p. 233-237.
26. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
27. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.
28. Ahlmann-Eltze, C. (2017). ggsignif: Significance Bars for 'ggplot2'. R package version 0.2.0. https://CRAN.R-project.org/package=ggsignif.
29. Harrell Jr, F. E., with contributions from Charles Dupont and many others. (2016). Hmisc: Harrell Miscellaneous. R package version 4.0-2. https://CRAN.R-project.org/package=Hmisc.
30. Gasteiger, E., et al., *Protein Identification and Analysis Tools on the ExPASy Server*, in *The Proteomics Protocols Handbook, Humana Press*, J.M. Walker, Editor. 2005, Humana Press.
31. Mayer, O.H., et al., *Characterization of pulmonary function in Duchenne Muscular Dystrophy.* Pediatr Pulmonol, 2015. **50**(5): p. 487-94.
32. Kerr, T.P., et al., *Long mutant dystrophins and variable phenotypes: evasion of nonsense-mediated decay?* Hum Genet, 2001. **109**(4): p. 402-7.

Appendix A. Analysis of genotype and clinical phenotype

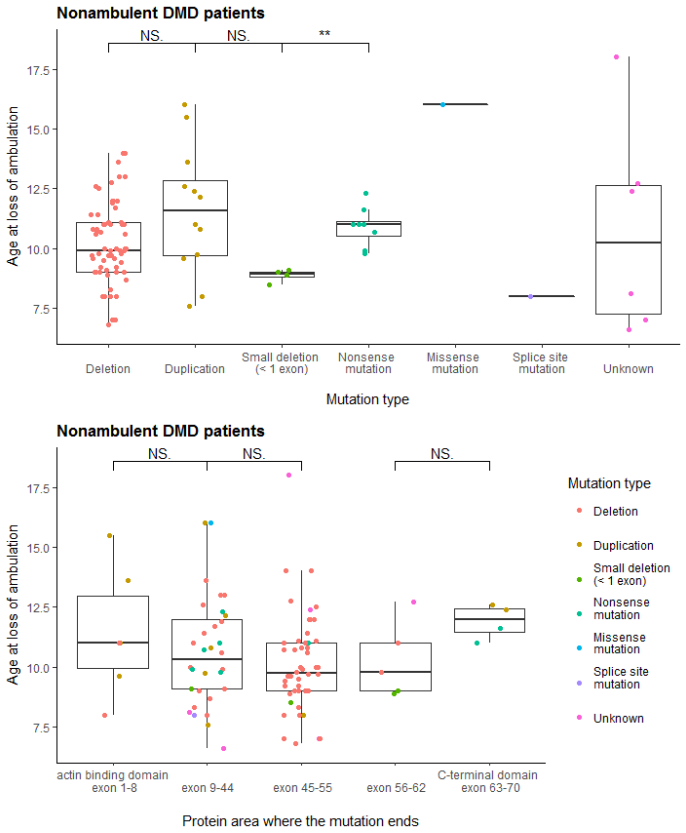

Figure 6. The effect of mutation type and mutation location on the age of complete loss of ambulation in nonambulent DMD patients from all three cohorts. Wilcoxon rank sum test was used to determine if two groups were significantly different from one another. N.S. = no significant difference, * = significance level p-value < 0.1, ** = significance level p-value < 0.01, *** = significance level p-value < 0.001.

Appendix B. Figures for antibody suspension bead array data form the UNEW cohort.

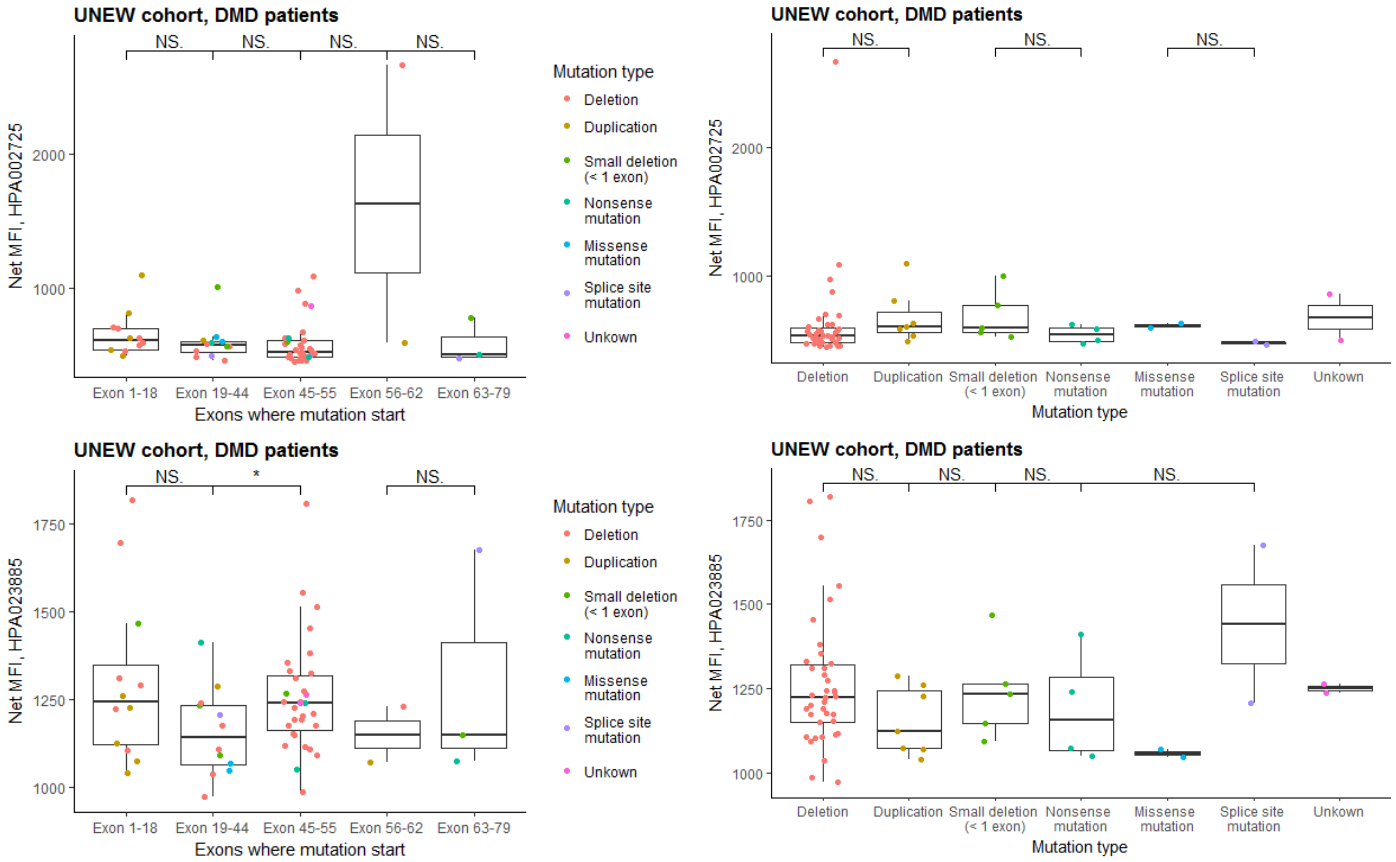

Figure 7. The effect of mutation type and mutation location on the levels of N-terminal (HPA002725) and C-terminal (HPA023885) dystrophin levels in serum for DMD patients in the UNEW cohort. Wilcoxon rank sum test was used to determine if two groups were significantly different from one another. N.S. = no significant difference, * = significance level p-value < 0.1, ** = significance level p-value < 0.01, *** = significance level p-value < 0.001.

Appendix C. Age dependency and FVC % of predicted dependency of antibody suspension bead array data from UNEW

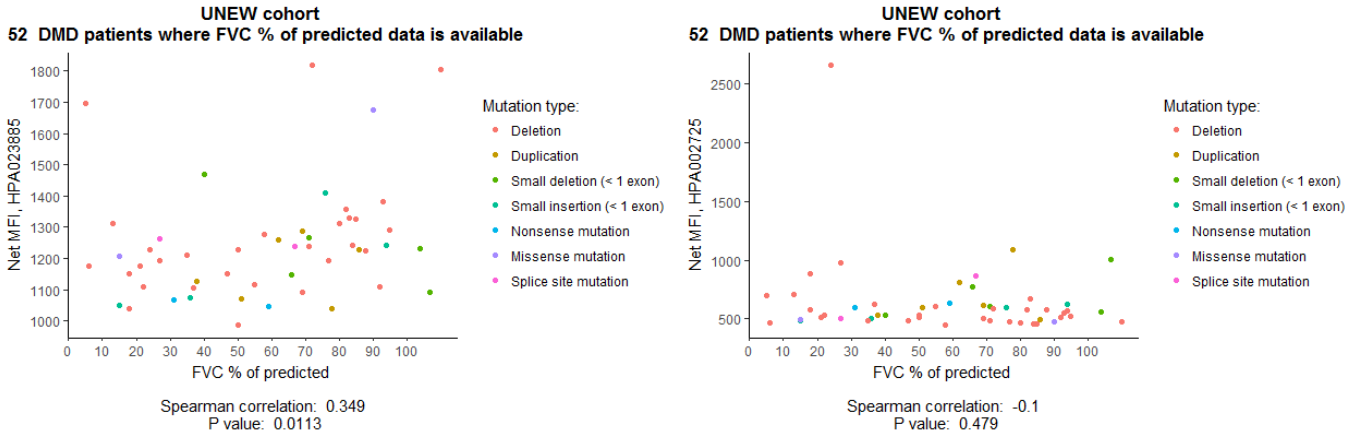

Figure 8. Analysis of correlation between the Net MFI of HPA023885 (left) and HPA002725 (right) in serum samples from 52 DMD patients in the UNEW cohort, and the FVC % of predicted value measured for these patients at the time of sample collection.

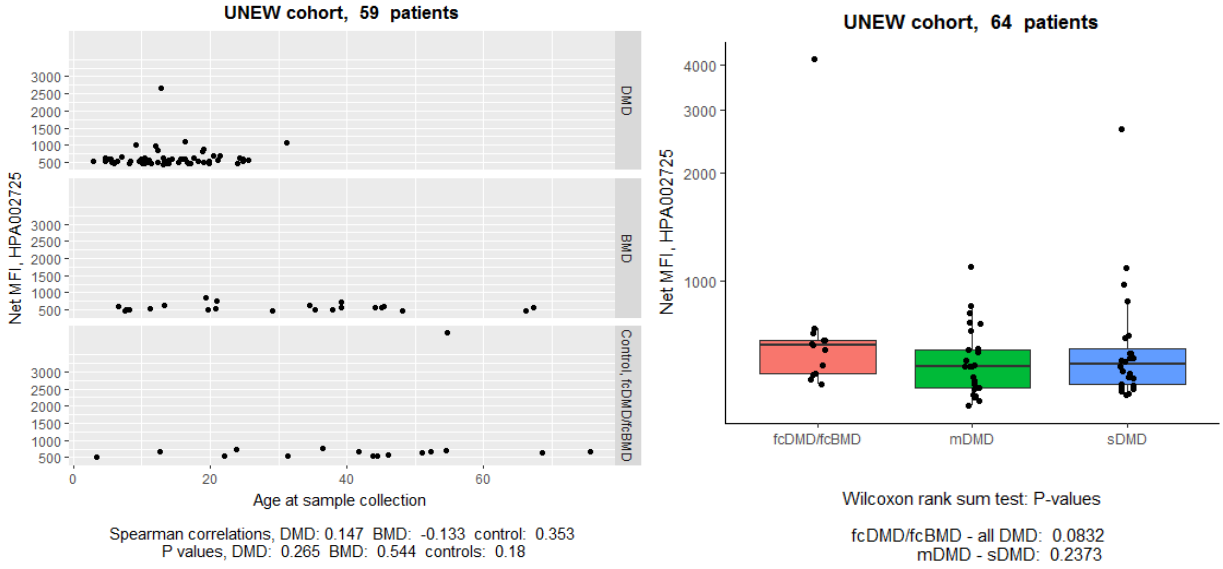

Figure 9. Left: The age dependency of Net MFI for the N-terminal dystrophin antibody, HPA002725, in serum samples from patients with DMD (top plot), BMD (middle plot) and female carrier control samples (bottom plot). **Right:** The differences in Net MFI for antibody HPA002725 between patients classified as having a mild DMD phenotype (mDMD, age of loss of ambulation > 13) and patients classified as having a severe phenotype (sDMD, age of loss of ambulation < 13), as well as control patients consisting of female carriers of either DMD or BMD (fcDMD/fcBMD).

Appendix D. Raw bead array data from LUMC cohort

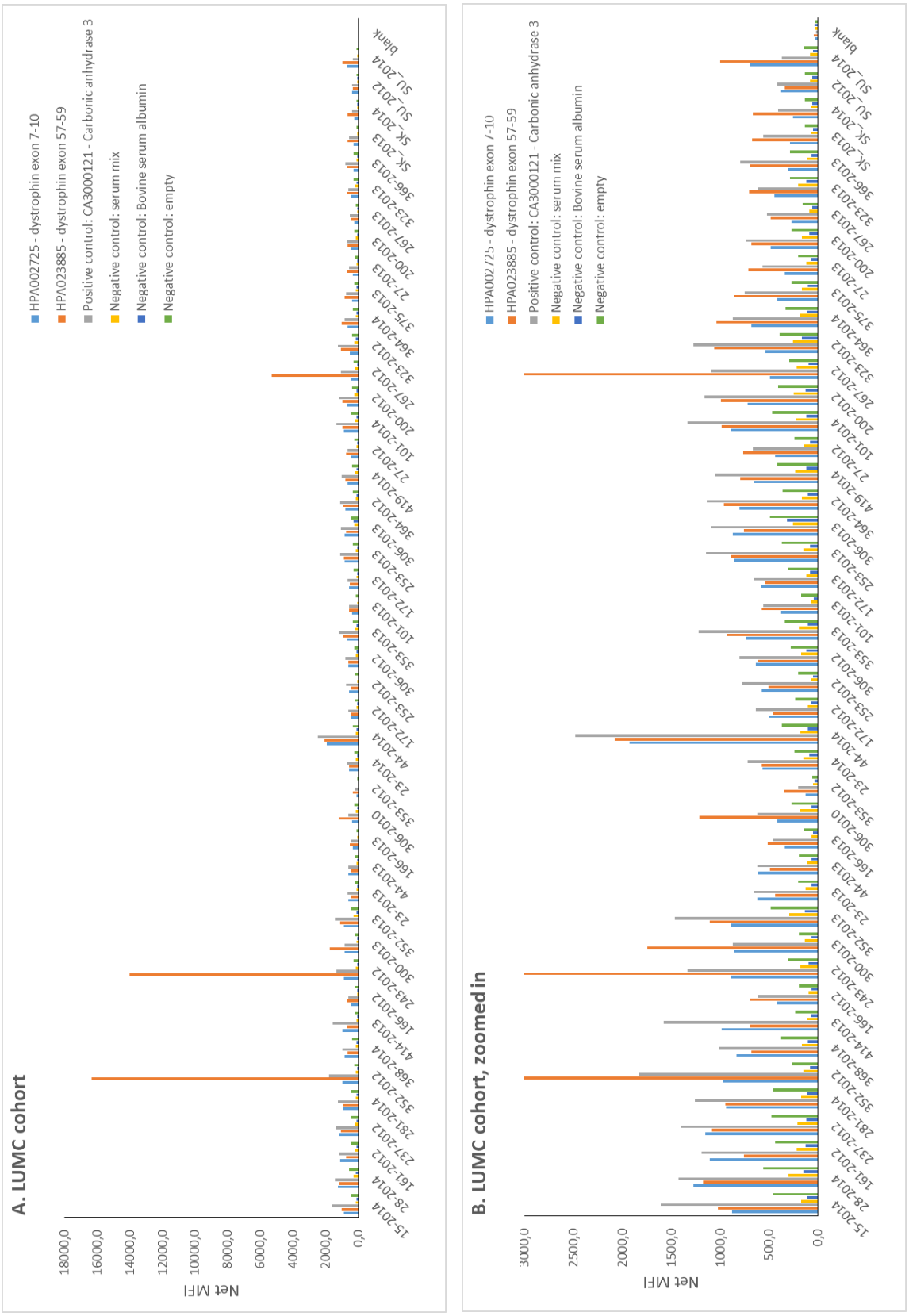

Figure 10. Box chart representing raw data net mean fluorescence intensity (Net MFI) from protein bead arraw analysis of the LUMC cohort.

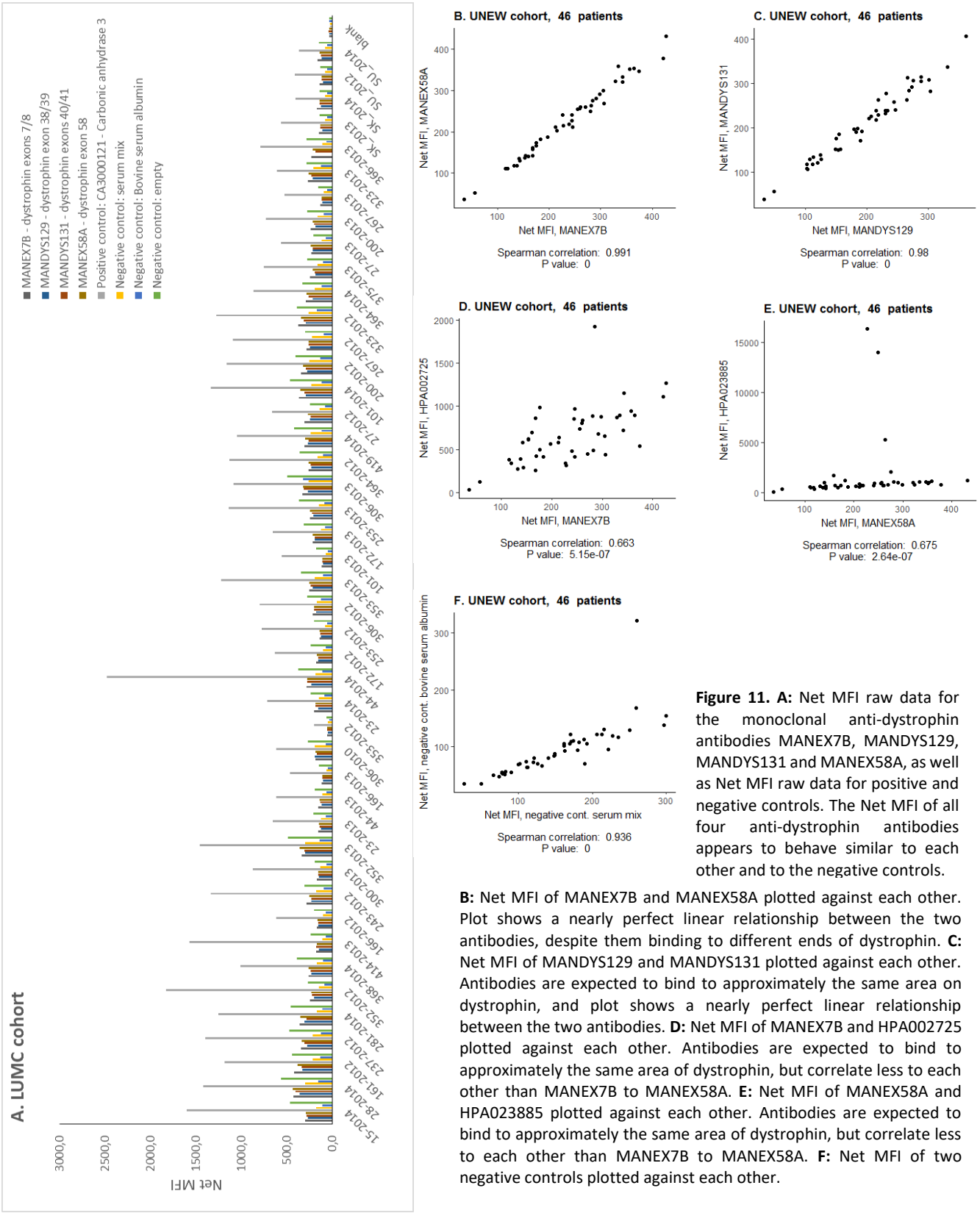

Figure 11. A: Net MFI raw data for the monoclonal anti-dystrophin antibodies MANEX7B, MANDYS129, MANDYS131 and MANEX58A, as well as Net MFI raw data for positive and negative controls. The Net MFI of all four anti-dystrophin antibodies appears to behave similar to each other and to the negative controls.

B: Net MFI of MANEX7B and MANEX58A plotted against each other. Plot shows a nearly perfect linear relationship between the two antibodies, despite them binding to different ends of dystrophin. **C:** Net MFI of MANDYS129 and MANDYS131 plotted against each other. Antibodies are expected to bind to approximately the same area on dystrophin, and plot shows a nearly perfect linear relationship between the two antibodies. **D:** Net MFI of MANEX7B and HPA002725 plotted against each other. Antibodies are expected to bind to approximately the same area of dystrophin, but correlate less to each other than MANEX7B to MANEX58A. **E:** Net MFI of MANEX58A and HPA023885 plotted against each other. Antibodies are expected to bind to approximately the same area of dystrophin, but correlate less to each other than MANEX7B to MANEX58A. **F:** Net MFI of two negative controls plotted against each other.

Appendix E. Net MFI of N-terminal and C-terminal antibody, comparison between DMD patients and healthy controls in LUMC cohort.

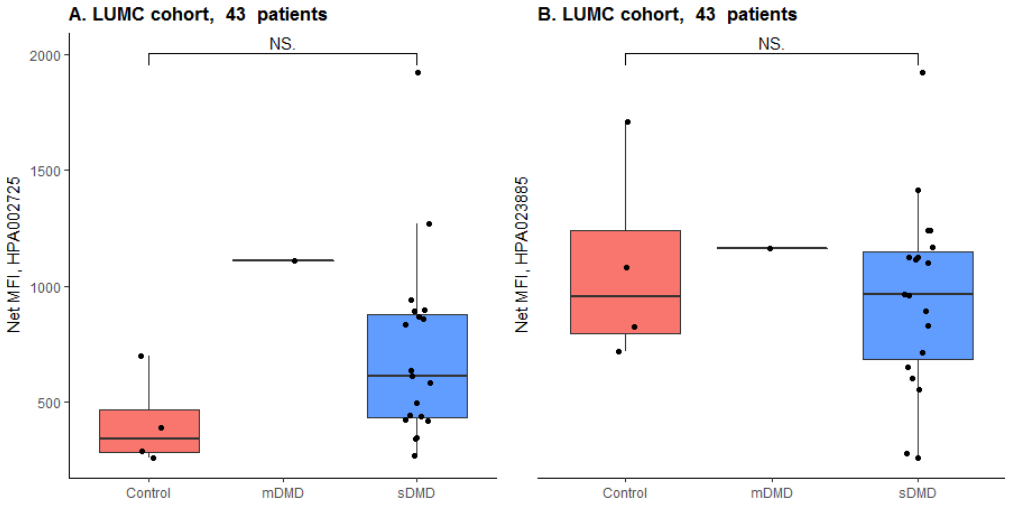

Figure 12. A. Net MFI of N-terminal antibody HPA002725 for severe DMD patients (sDMD), mild DMD patients (mDMD) and healthy controls. B. Net MFI of C-terminal antibody HPA023885 for severe DMD patients (sDMD), mild DMD patients (mDMD) and healthy controls. In both plots, disease severity is defined as patients who lost ambulation before age 13 (sDMD) or after (mDMD). Wilcoxon rank sum test was used to determine if two groups were significantly different from one another. N.S. = no significant difference, * = significance level p-value < 0.1, ** = significance level p-value < 0.01, *** = significance level p-value < 0.001.

Appendix F. Theoretical molecular weights for dystrophin splice variants

Table 5. All predicted protein coding splice variants enlisted in Ensembl.org as of 2017-05-15 [14], with molecular weights calculated using the ExPASy bioinformatics tool "Compute pI/Mw" [30].

Isoform name	Theoretical Mw (ExPASy), Da	First exon compared to DMD-001	Last exon compared to DMD-001	Antibody HPA002725 N-terminal	Antibody HPA023885 C-terminal	Transcript variant name(s)	RefSeq
DMD-001	426 777.67	1	79	YES	YES	Dp427c	NM_000109 NM_004006 NP_000100 NP_003997
DMD-006	426 028.78	1	79	YES	YES	Dp427p	NM_004009 NM_004010 NP_004000 NP_004001
DMD-203	141 399.08	45	79	-	YES	Dp260_1 Dp260_2 Dp140	NM_004011 NM_004012 NM_004013 NP_004002 NP_004003 NP_004004
DMD-204	128 977.62	45	79	-	YES	Dp140c Dp140bc	NM_004020 NM_004023 NP_004011 NP_004014
DMD-201	109 969.57	56	79	-	YES	Dp116	NM_004014 NP_004005
DMD-013	72 190.69	63	79	-	-	Dp71b	NM_004016 NP_004007
DMD-015	70 374.77	63	79	-	-	Dp71 Dp71a	NM_004015 NM_004017 NP_004006 NP_004008
DMD-014	70 750.03	63	79	-	-	Dp71ab	NM_004018 NP_004009
DMD-206	426 847.91	1	79	YES	YES		-
DMD-205	426 099.02	1	79	YES	YES		-
DMD-202	141 774.34	45	79	-	YES	Dp140ab	NM_004022 NP_004013
DMD-017	159 362.43	48	79	-	YES		-
DMD-016	143 215.01	45	79	-	YES	Dp140b	NM_004021 NP_004012
DMD-002	88 799.87	1	18	YES	-		-
DMD-019	59 769.23	63	79	-	-	Dp40	NM_004019 NP_004010
DMD-020	27 390.28	67	74	-	-		-
DMD-024	16 098.20	38	40	-	-		-
DMD-009	14 524.64	3	5	-	-		-
DMD-004	12 033.74	2	36	-	-		-
DMD-005	5 263.86	2	36	-	-		-