

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Biomedicina Comparata e Alimentazione

---

DOTTORATO DI RICERCA IN : SCIENZE VETERINARIE

CICLO XXVII

# PATHOGENETIC MECHANISMS OF COLLAGEN TYPE VI-RELATED MUSCULAR DYSTROPHIES

**Coordinatore :** Ch.mo Prof. GIANFRANCO GABAI

**Supervisore :** Ch.mo Prof. LUCA ARESU

**Dottorando** : ELEONORA GUADAGNIN

**Matricola:** 1040803

*Così diventò tutto piccolo*
*anche le notti là in America*
*Ti volti e vedi la tua vita*
*come la bianca scia di un'elica*
*(Caruso, Lucio Dalla, 1986)*

# TABLE OF CONTENT


SUMMARY 1

SOMMARIO 3

1 - INTRODUCTION 6

COLLAGEN TYPE VI-RELATED MUSCULAR DYSTROPHIES 6

1.1 - CLINICAL FEATURES 6

1.2 - ETIOLOGY 8

1.3 - PATHOGENESIS 9

1.3.1 - Collagen type VI Structure 9

1.3.2 - Collagen type VI Function 10

1.3.3 - Pathophysiology 10

2 - AIM OF THE STUDY 14

3 - MATERIALS AND METHODS 15

3.1 - Participants 15

3.2 - Histology and images acquisition 15

3.3 - Automated image analysis 15

3.4 - RNA Isolation and Quality Assessment 18

3.5 - Microarray Analysis 19

3.6 - RNA-Seq Analysis 19

3.7 - qRT-PCR expression validation 21

3.8 - Immunoblotting 21

4 - RESULTS 25

4.1 - Morphological characterization 25

4.2 - Differential gene expression analysis 27

4.3 - Differentially expressed splice isoforms analysis 33

4.4 - Upstream Regulators 35

4.5 - Validation of differentially expressed genes by qRT-PCR 37

4.6 - Validation of selected protein targets 39

5 - DISCUSSION 41

6 - BIBLIOGRAPHY 47

## SUMMARY

The overall focus of thesis is determining key disease-driving molecular pathways involved in the progression of a specific subtype of neuromuscular diseases, the collagen type VI-related muscular dystrophies (COL6RD). These kinds of muscular disorder encompass a wide spectrum of diseases, ranging from the severe Ullrich muscular dystrophy (UCMD) to the mild Bethlem myopathy (BM), characterized by a plethora of clinical manifestations. Both UCMD and BM usually arise at birth or early childhood with hypotonia and progressive muscle weakness, joint contractures and hyperlaxity of the distal joints. In UCMD, ambulation is achieved with a delay of 2 years compared to healthy children, but it is eventually lost. In BM, ambulation is generally maintained and only few affected individuals require wheel chair. Genetically, UCMD and BM are caused by mutations in the genes encoding for the three α-chains of the collagen type VI acting in a recessive or dominant way. Morphologically, the muscle biopsies of UCMD and BM patients are characterized by variable degree of atrophy, fibrosis, and regeneration and degeneration of the muscle fibers.

In the present work, we systematically reviewed and classified a cohort of 22 COL6RD patient muscle biopsies and 14 age-matched control muscle biopsies according to basic histological features, such as muscle fibers atrophy and fibrosis. This classification was achieved through an innovative automatic image analysis algorithm and generated three groups, COL6RD-1, COL6RD-2 and COL6RD-3, of increasing histological severity. Microarray-based gene expression profiling and RNA-Seq of the same cohort produced gene expression fingerprint that suggested a strong segregation of COL6RD muscle biopsies from the control muscle biopsies, but did not clearly correlated with the histological grouping established among the patients. *In silico* analysis of the gene expression results revealed TGFβ1 pathway as the major regulator of the COL6RD histological changes and, most likely, of the underlying disease pathophysiology. TGFβ1 involvement was further confirmed by upstream regulator analysis, validation of downstream gene targets by qRT-PCR, and protein targets by western blot and immunofluorescence. Moreover, we also found a downregulation of skeletal muscle specific genes.

RNA-Seq data also highlighted a deeper level of regulation of the gene expression carried over by alternative splicing. We identified 43 genes having more than one splice isoform whose expression level was significantly different between COL6RD muscle biopsies and controls, with DECORIN being the one displaying the highest number of alternatively spliced isoforms.

Overall, this study suggests that different histological stages of COL6-RD are characterized by the same pattern of gene expression, involving predominantly the upregulation of matricellular components, even when the muscle involvement is not evident yet, and downregulation of muscle specific genes. Our results also suggest that TGFβ1 acts as a major driver of the pathological changes at the skeletal muscle level. Moreover, muscle biopsies from COL6-RD patients are characterized by an increase in alternative spliced isoforms of extracellular matrix-related genes that might contribute to the pathogenesis of the disease.

# SOMMARIO

L'obiettivo di questa tesi è determinare i meccanismi molecolari coinvolti nella progressione di un sottogruppo specifico di malattie neuromuscolari: le distrofie muscolari congenite causate da deficit di collagene di tipo VI. Questo gruppo di distrofie comprende un ampio spettro di disordini caratterizzati da un ampio spettro di manifestazioni cliniche: dalla più grave distrofia muscolare congenita di Ulrich (UCMD) alla più moderata miopatia di Bethlem" (BM). Sia UCMD che BM si manifestano alla nascita o nel prima infanzia con ipotonia e debolezza muscolare progressiva, contrazioni delle articolazioni prossimale e iperelasticità delle articolazioni distali. Individui affetti da UCMD iniziano a deambulare con un ritardo di 2 anni rispetto ad individui sani, e la capacità deambulatoria viene poi persa nuovamente. Individui affetti da BM, invece, conservano la capacità di ambulare e solo pochi necessitano di sussidi. Dal punto di vista genetico, UCMD e BM sono causate da mutazioni che possono colpire uno o più geni codificanti le tre molecole del collagene di tipo VI. Tali mutazioni possono agire in modo dominante o recessivo. Morfologicamente le biopsie muscolari di pazienti affetti da UCMD e BM sono caratterizzate da atrofia muscolare, fibrosi, rigenerazione e degenerazione delle fibre muscolare, evidenziata dalla presenza di aree necrotiche.

Nel presente lavoro, abbiamo effettuato un'analisi sistematica di 22 biopsie muscolari ottenute da pazienti affetti da COL6RD e 14 biopsie muscolari ottenuti da controlli sani di età corrispondente, che sono state suddivise in base al grado di atrofia muscolare e fibrosi. Questa classificazione è stata effettuata usando un algoritmo originale per l'analisi automatica di preparazioni istologiche, ed ha generato 3 gruppi, COL6RD-1, COL6RD-2 and COL6RD-3, di gravità istologica crescente. Le stesse biopsie sono state sottoposte a microarray e RNA-Seq ed hanno rivelato andamenti di espressione genica in forte correlazione con i gruppi definiti secondo secondo i preparati istologici, sottolineando la chiara divisione tra le biopsie muscolari di individui affetti da COL6RD e controlli. L'analisi *in silico* dell'espressione genica ha successivamente rivelato TGFβ1 come il maggior responsabile dei cambiamenti istologici in COL6RD e, verosimilmente, dei processi patofisiologici sottostanti. Il coinvolgimento di TGFB1 è stato ulteriormente confermato dall' *upstream regulator analysis*, validazione di geni target downstream tramite qRT-PCR, e di proteine effettrici tramite western blot e immunofluorescenza. I dati ottenuti tramite RNA-Seq hanno rivelato un ulteriore livello di regolazione dell'espressione genica attuato tramite splicing alternativo. Abbiamo identificato 43 geni aventi più di una variante la cui espressione è significativamente diversa tra biopsie muscolari di COL6RD e controlli sani. Nello specifico, Decorin è il gene con il maggior numero di isoforme originate da splicing alternativo.

Nel complesso, questo studio suggerisce che i diversi stadi istologici di COL6RD sono caratterizzati dallo stesso profilo di espressione genica, che coinvolge prevalentemente l'aumento di componenti della matrice extracellulare, anche in mancanza di evidenze istologiche di fibrosi, e la diminuzione di prodotti di geni muscolo-specifici.

I nostri risultati suggeriscono anche che ci siano pathway molecolari TGFβ1-dipendenti che agiscono da principali regolatori dei cambiamenti patologici a livello del muscolo scheletrico. Inoltre, le biopsie muscolari provenienti da pazienti affetti da COL6RD sono caratterizzate da un aumento delle isocrone geniche generate da splicing alternativo, soprattutto per geni coinvolti nella produzione e nel mantenimento della matrice extracellulare.

# 1 - INTRODUCTION

# COLLAGEN TYPE VI-RELATED MUSCULAR DYSTROPHIES

## 1.1 - CLINICAL FEATURES

Collagen type VI-related muscular dystrophies (COL6RD) are a clinically heterogeneous group of neuromuscular disorders characterized by muscle weakness and hypotonia, proximal joint contractures and distal joint hyperlaxity noted at birth or early infancy. COL6RD can be ideally represented by a spectrum of conditions, with Ullrich congenital muscular dystrophy being at the most severe end, and Bethlem myopathy being the milder end.

Ullrich Congenital Muscular Dystrophy (UCMD) was originally described by Otto Ullrich in 1930 as a congenital muscular dystrophy that he called "scleroatonic muscular dystrophy". In the late '70s the same condition was reported in four Japanese individuals. Two of them were siblings from consanguineous parents with clinical history of muscular disorders. The other two patients were unrelated. All of them were described as floppy at birth. They were never able to crawl or walk, or they acquired ability to walk later in infancy but they lost it very soon. Muscle weakness was slowly progressive, with the distal muscles being preferentially involved. Serum creatine kinase (CK) levels were normal. Electromyographic examination showed a myopathic pattern. Muscle biopsies from the four patients showed dystrophic features: variation in muscle fiber size, adipose tissue replacement, centrally located nuclei. All four patients were very intelligent (Furukawa & Toyokura, 1977).

UCMD was also reported in two Italian siblings from healthy, non consanguineous parents, with relatively late onset. Both patients were healthy at birth, but they started to develop progressive muscle weakness, wobbly gait, difficulty climbing stairs, and distal joint hyperlaxity around 3 years of age. At the age of 7 years they started to develop major proximal joints contractures and rigid spine. They both showed progressive decreased respiratory capacity. Muscle biopsy of the deltoid muscle showed variation in fiber size, reduction of the muscular component that was replaced by connective tissue, prominent mostly in the endomysium and perimysium, and fat tissue. CK levels were normal or mildly elevated. Intellectual development was normal for both siblings (Ricci et al., 1988).

Bethlem Myopathy (BM) was originally described in the late '70 in 3 Dutch families in which 28 patients were affected by a benign myopathy with autosomal dominant inheritance. The onset of the disease was noted early in infancy, approximately at the age of five, although the

patients were not born as floppy babies. Developmental milestones were achieved normally. At young age, many patients tended to fall down easily, they could not run as fast as their peers, and they had difficulty at gymnastics with jumping and skipping. After this early manifestation of symptoms, the progression of the disease was otherwise very slow. The benign character of the disorder is testified by the fact that life expectancy was not shortened. The patients presented with a generalized muscle weakness and muscle atrophy in adulthood. Proximal muscles were more affected than distal, and the extensor muscles more than the flexors. Contractures of the elbows, ankles and of the interphalangeal joints of the last four fingers were also present. Muscle biopsies revealed marked variation in fiber size and marked increased in fatty tissue. CK levels were normal or only mildly elevated (Bethlem & Wijngaarden, 1976). Many case reports followed afterwards describing the same benign autosomal dominant neuromuscular disorder characterized by early onset but very slow progression, elbows and ankles contractures and mild non-specific myopathic changes on histology (Arts, Bethlem, & Volkers, 1978; Leyten et al., 1986; Merlini, Morandi, Granata, & Ballestrazzi, 1994; Mohire et al., 1988).

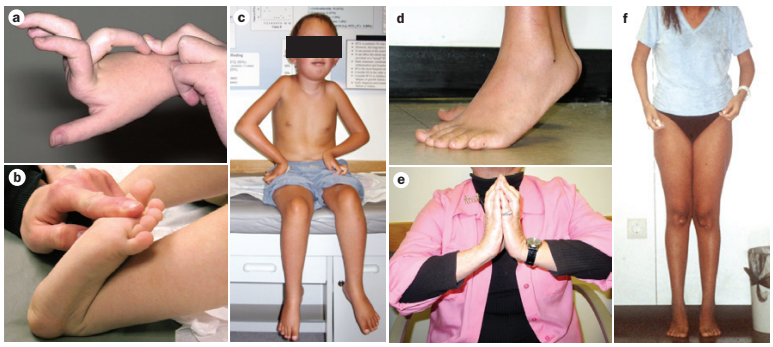



**Figure 1: Joint laxity and progressive joint contractures in patients with UCMD and BM.** (a,b) Typical distal joint hyperlaxity, particularly in the finger, in patient with UCMD. (c) Contractures of the major joint (shoulders, elbows, and knees) in a patient with UCMD. (d,e) Contractures of the Achilles tendon and finger flexor in a patient with BM. (f) Contractures of shoulder and elbow joints in a patient with BM (Bonnemann, 2011)

Studies based on muscle magnetic resonance imaging (MRI) revealed a very specific pattern of muscle involvement concerning both Bethlem myopathy and Ullrich congenital muscular dystrophy patients (Mercuri et al., 2002; Mercuri et al., 2003; Mercuri et al., 2005). Severe clinical phenotype are usually associated with severe and diffuse involvement of the calf muscle and of the thigh muscles, with relative sparing of sartorius, gracilis, adductor longus and rectus. Vastus lateralis is peculiar in that it shows the typical pattern of concentric involvement, while the rectus femoris is characterized by fatty tissue infiltration under the central fascia, also known as central shadow. Moderate clinical manifestations reflect a moderate muscle involvement, with only minimal changes in the mildest cases.

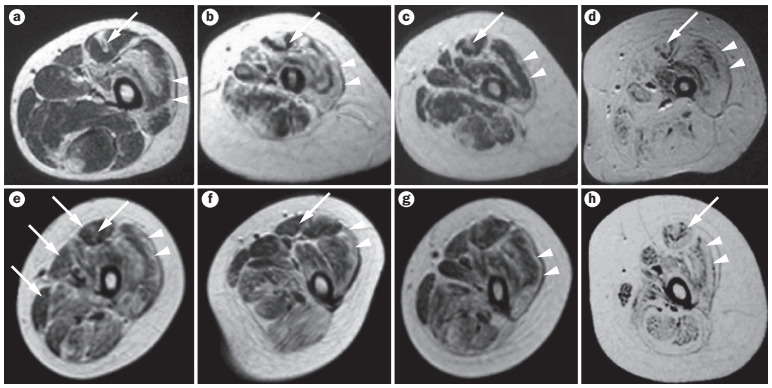


**Figure 2: Muscle MRI of patients with UCMD and BM.** (a-d) BM (e-h) UCMD. T1-weighted MRI scans show areas of fatty tissue replacement in the typical "outside-in" pattern. The involved areas are the center of the rectum femoris (long arrows) and the vastus lateralis (arrowheads) (Bonnemann, 2011)

### 1.2 - ETIOLOGY

COL6A1 and COL6A2 are both located on chromosome 21q22.3. They code for the α1(VI) and the α2(VI). COL6A3 is located on 2q37 and codes for the α3(VI) chain. COL6A1 and COL6A2 are located in a head-to-tail fashion, separated by an intermediate region of approximately 93kb. COL6A3 is regulated independently from COL6A1 and COL6A2(Chu et al., 1989; Lamande, Sigalas, et al., 1998).

Collagen type VI-related muscular dystrophies (COL6RD) are caused by mutations in any of three genes encoding the subunits of collagen type VI. Both Ullrich congenital muscular dystrophy and Bethlem myopathy are inherited as autosomal dominant or recessive conditions. However, de novo dominant-acting mutations account for the majority of UCMD cases, and autosomal recessive BM are a very rare event (Baker et al., 2005; Foley et al., 2009; Gualandi et al., 2009). A variety of mutations at the genomic or coding DNA level are known to cause COL6RD.

The most common events that lead to UCMD and BM are heterozygous missense mutations on COL6A1, COL6A2 or COL6A3 genes that cause amminoacid substitutions affecting different domains of the α1(VI), α2(VI), and α3(VI) subunits of the collagen VI protein. The most common effects are changes in the Gly residue of the invariable Gly-X-Y motif at the triple helical domain or at the N-terminal or C-terminal globular domains of the α1(VI), α2(VI), and α3(VI) (Demir et al., 2002; Jobsis et al., 1996; Pan et al., 1998; Pepe, Bertini, et al., 1999; Scacheri et al., 2002; Vanegas et al., 2002). Mutations causing premature termination and in-frame deletions can also affect the mRNA maturation, leading to skipping of exons or truncation of the mRNA harboring the mutation, preventing the biosynthesis of fully functional α-chains (Lamande, Bateman, et al., 1998; Lampe et al., 2008; Pepe, Giusti, et al., 1999; Pepe et al., 2006; Vanegas et al., 2002). Nonsense-mediated decay of the mRNA carrying the mutations is also a mechanism commonly found in UCMD and BM patients(Lampe et al., 2008; Lucarini et al., 2005; R. Z. Zhang et al., 2002).

Three additional collagen VI genes were identified and characterized, COL6A4, COL6A5, and COL6A6 (Fitzgerald, Rich, Zhou, & Hansen, 2008; Gara et al., 2008). These genes are clustered at a single genomic locus on chromosome 3, 3p24.3, which is very conserved in mammals. The products of these genes are three α-chains, α4(VI), α5(VI), and α6(VI). In humans, COL6A6 is widely expressed, but COL6A5 is more limited. COL6A4, instead, is not expressed in humans due to a disruption of the gene occurred during the speciation process. In mouse, *Col6a5* and *Col6a6* are expressed in a wide range of tissues. *Col6a4* is limited to few tissues, such as uterus and ovary.

Interestingly, *in vitro* studies have shown that α5(VI) and α6(VI) chains are not competent to assemble with the three main α-chains of collagen VI, α1(VI), α2(VI), and α3(VI). However, to date, there are no reported mutations in the COL6A5 and COL6A6 genes in patients affected by COL6RD, and their involvement in human diseases remains unknown.

## 1.3 - PATHOGENESIS

### 1.3.1 - Collagen type VI Structure

Collagen type VI protein is composed by three polypeptides, α1(VI), α2(VI), and α3(VI) chains, that assemble in stoichiometric proportions.

COL6A1 consists of a total of 36 exons: 7 code for the globular N-terminal domain of α1(VI) chain, one fusion exon coding for the final part of the N-terminal domain and the initial part of the triple helical domain, 19 coding uniquely for the triple helical domain, and 9 coding for the C-terminal domain. α1(VI) C-terminal domain can be further subdivided in two subdomains: C1 and C2 (Trikka, Davis, Lapenta, Brahe, & Kessling, 1997).

COL6A2 consists of a total of 28 exons: 4 exons code for the N-terminal domain of the α2(VI) chain, one fusion exon coding for the final part of the N-terminal domain and the initial part of the triple helical domain, 18 coding uniquely for the triple helical domain, and 5 coding for the C-terminal domain. The C-terminal domain of α2(VI) is encoded by two exons that undergo alternative splicing and originate two different C-termini (Saitta, Timpl, & Chu, 1992).

COL6A3 consists of a total of 44 exons: 12 exons code for the N-terminal domain of the α3(VI) chain, one fusion exon coding for the final part of the N-terminal domain and the initial part of the triple helical domain, 18 coding uniquely for the triple helical domain, and 12 coding for the C-terminal domain. Exons from the N-terminal and C-terminal domains undergo tissue-specific alternative splicing originating multiple transcripts (Stokes, Saitta, Timpl, & Chu, 1991).

The triple helical region of each α-chain is characterized by continuous Gly-X-Y repeats, two key cysteine residues, and abundance of RGD (Arg-Gly-Asp) tripeptide sequence, important for cell binding activity. The triple helical domain of α1(VI) is coded by 18 small exons and contains three RGD sequences. The triple helical region of α2(VI) is coded by 19 exons and contains five RGD sequences (Saitta et al., 1991).

Assembling of the fully functional collagen type VI protein is carried over intracellularly and extracellularly. Inside the cell, the three α-chains assemble to form monomers. Two monomers are laterally aligned in an antiparallel fashion, with the globular domain interacting with the adjacent helix, to form disulfide-bonded dimers. Two dimers form tetramers that are stabilized by disulfide bonds formed within the triple helical domains. The tetramers are then exported into the extracellular space, where they assemble in a end-to-end fashion to form beaded-looking microfibrils (Bruns, 1984; Kuo, Keene, & Glanville, 1995)(Chu et al., 1987; Engel et al., 1985).

### 1.3.2 - Collagen type VI Function

Collagen type VI is a non fibrillar collagen expressed in many connective tissues and that is implicated in the extracellular matrix organization. In skeletal muscle, collagen type VI is produced by the interstitial fibroblasts and it is thought to anchor the basement membrane in the skeletal muscle by interacting with collagen IV, a major component of the basal lamina (Kuo, Maslen, Keene, & Glanville, 1997; Zou, Zhang, Sabatelli, Chu, & Bonnemann, 2008).

### 1.3.3 - Pathophysiology

COL6RD affects two major bodily compartments: extracellular matrix and skeletal muscle. With regards to the extracellular matrix, the collagen VI protein can be reduced, mislocalized, or completely absent. We have previously described how apparently similar mutations in the three collagen VI genes, COL6A1, COL6A2, and COL6A3, cause both mild Bethlem myopathy phenotype and severe Ullrich congenital muscular dystrophy phenotype. The outcome depends greatly on the position of the mutation, the extension of the involvement of the collagen VI genes, and the consequences that these have on the complex biosynthesis of the collagen VI monomers, dimers, tetramers, and ultimately of the microfibrils.

Generally speaking, mutations that affect the globular C-terminal domain of any of the α-chains have more severe consequences than mutations affecting the N-terminal domain. Mutations affecting the triple helical domain of any of the α-chains can have more severe or milder consequences according to their relative position respect to the C-terminal or N-terminal domain,

and to the amminoacid sequence that they affect. Moreover, the extend of the downstream effects depends also on the ability of the mutated α-chains to be incorporated into monomers, dimers, tetramers and microfibril of collagen VI, and the grade of dominant-negative effect that they exert on the assembly of monomers, dimers, and tetramers.

*In frame deletions*

Heterozygous mutations causing in-frame deletions and resulting in exon skipping have been reported for both Bethlem myopathy and Ullrich congenital muscular dystrophy. In studies previously published studies, the exons skipping involved exon 14 of the α1(VI), exon 11 of the α2(VI), exon 15 of the α3(VI), all of them located within the triple helical domain of α1(VI), α2(VI) and α3(VI) (Baker et al., 2007; Lamande, Shields, Kornberg, Shield, & Bateman, 1999; Pan et al., 2003; Pepe, Giusti, et al., 1999)

The skipping of any exon in the coding DNA of COL6A1, COL6A2, and COL6A3 would lead to the synthesis of internally deleted α-chains with possibly compromised ability to form monomers, dimers, and tetramers, depending on the quality of the defects. If the α-chains were severely structurally compromised or the deletion affected critical amino acid residues necessary for the stability of the collagen VI protein, for example cysteine residues necessary for the formation of disulfide bonds between monomers and dimers, the monomers would not associate into more complex structures, and the mutated α-chain would not be secreted. Because the mutation appears in homozygosity, and stable collagen VI monomers forms with the stoichiometry 1 α1(VI): 1 α2(VI): 1 α3(VI), the result is the strongly reduced amount of structurally normal collagen VI in the extracellular matrix.

Mutations causing in-frame deletions exerting a strong dominant-negative effect and compromising the intracellular assembly of dimers, tetramers, and extracellular microfibrils, generally result in a severe deficiency of collagen VI in the extracellular matrix that can be associated with a severe Ullrich congenital muscular dystrophy phenotype. If the mutations was located proximal to the N-terminal end of triple helical domain, the patient phenotype would likely resemble the mild Bethlem myopathy.

*Glycine substitutions*

A class of mutations that have been associated with both Bethlem myopathy and Ullrich congenital muscular dystrophy is Glycine (Gly) substitutions that disrupt the invariable Gly-X-Y motif of the triple helical domain (Giusti et al., 2005; Jobsis et al., 1996; Lamande et al., 2002; Pepe, Bertini, et al., 1999; Sasaki et al., 2000; Scacheri et al., 2002). Gly substitutions in the conserved Gly-X-Y motif are the most commonly identified mutations in COL6RD, accounting for roughly one third of all the pathogenic mutations in COL6A1, COL6A2, and COL6A3. The majority of Gly substitutions in the triple helical domain are *de novo,* dominantly-acting mutations. However, recessive cases have been reported. Interestingly, the dominant negative mutations are generally more frequently found in the N-terminal end of the α-chain triple helix, while the recessive mutations are near the C-terminal end of the triple helical domain. Structural analyses have suggested that the critical region important for microfibril formation is the segment included between the Gly-X-Y triplets 10-17, and mutations occurring within the critical region are the leading cause of more severe clinical phenotype. The seventeenth Gly-X-Y triplet is close to a cysteine residue on the α3(VI) chain that is essential for tetramers stabilization via disulfide bonds. Mutations occurring outside of this critical region, toward the N-terminal end of the triple helix, do not prevent the formation of tetramers, although very unstable. The resulting collagen VI deposited in the extracellular matrix would contain structural abnormalities that would reduce the efficiency in the end-to end association to form microfibrils. Patients presenting these kind of mutations also present with a wide spectrum of clinical severity, but generally milder phenotypes likely due to less compromised microfibril assembly (Butterfield et al., 2013).

*Nonsense-mediated decay of the mutant mRNA*

Nonsense-mediated decay has been described both in Bethlem myopathy and Ullrich congenital muscular dystrophy (Lamande, Bateman, et al., 1998; Lucarini et al., 2005; R. Z. Zhang et al., 2002). It is due to insertions and point mutations that are usually the cause of a frameshift that ultimately leads to a premature stop codon in the mRNA of any of the collagen VI genes (COL6A1, COL6A2, and COL6A3). Premature stop codons in the final exon of a gene do not usually impact the mRNA stability, therefore the result is a truncated protein. The result of a premature stop codon in any but the final exon of the gene causes, instead, a strong reduction in the stability of the mRNA that leads to the absence of the mutant mRNA. The effect of a homozygous mutation that cause s nonsense-mediated decay of the mRNA is the complete or extremely severe reduction of the amount of collagen type VI produced, with patients carrying these mutations showing a Ullrich congenital

muscular dystrophy clinical phenotype. Heterozygous mutations affecting mRNA stability, instead, causes protein haploinsufficiency which results in Bethlem myopathy.

*Skeletal muscle ultrastructural changes*

Defects in collagen type VI are also associated with skeletal muscle ultrastructural changes, in Col6a1$^{-/-}$ mice and COL6RD patients (Irwin et al., 2003; Merlini et al., 2008). The structural alterations present in Col6a1$^{-/-}$ mice include damaged and irregularly shaped muscle fibers, defects in the shape and function of mitochondrial, and altered structure of the sarcoplasmic reticulum (SR). Mitochondria are known to respond to pathological conditions increasing the inner membrane permeability via the opening of the permeability transition pore (PTP). Opening of the PTP, in turn, alters the handling of the intracellular $Ca^{2+}$. SR and PTP in Col6a1$^{-/-}$ are involved in the abnormal $Ca^{2+}$ response when handling the excitation-contraction coupling. Moreover, Col6a1$^{-/-}$ muscle fibers have more apoptotic nuclei compared to normal control muscles. This was also proved in COL6RD patients. The mitochondrial dysfunction in absence of collagen VI might be ascribed to the abnormal engagement of integrins, probable binding partners of collagen VI in the extracellular matrix. Mishandling of $Ca^{2+}$ by the SR caused by the mitochondrial dysfunctions would cause further increase of the PTP opening, setting in motion the pathomechanism that eventually lead to structural defects of both mitochondria and SR.

More recently, deregulation of the autophagic process has been associated to the pathogenesis of the dystrophic phenotype of Col6a1$^{-/-}$ (Grumati et al., 2010). Impaired autophagy in Col6a1$^{-/-}$ would cause accumulation of abnormal and dysfunctional organelles and apoptosis of the muscle fibers caused by the increased production of reactive oxygen species. Forced induction of autophagy is able to prevent this pathological mechanism and protect Col6a1$^{-/-}$ myofibers from degeneration.

## 2 - AIM OF THE STUDY

Transcriptome-wide analysis using microarray and RNA-Sequencing (RNA-Seq) based approaches have been widely used to help elucidate molecular mechanisms and pathways active in healthy and diseased skeletal muscle, both in animal models and in human samples (Bakay, Zhao, Chen, & Hoffman, 2002; Palstra et al., 2013; Porter et al., 2002; Zhou, Dimachkie, Xiong, Tan, & Arnett, 2004). Compared to microarray, which is a well established and relatively inexpensive method allowing for comparative gene expression analysis, RNA-Seq allows to more quantitatively assess individual transcripts including splice variants within the transcriptome (Ozsolak & Milos, 2011). While microarray is based on hybridization of samples to a predetermined number of probes designed on a preexisting database, RNA-Seq allows gene expression quantification, and discovery of new gene and transcript variants only by directly interrogating the genome of reference without the possible limitation of being restricted to just transcripts represented in a single array. Furthermore, RNA-Seq performed at sufficient coverage has a broader dynamic range to detect and quantify transcripts that are rare and expressed below the threshold of detection of microarray, while also avoiding the interference from non-specific background signal resulting from cross-hybridization in microarray. However, RNA-Seq is computationally very demanding and necessitates complex algorithms and bioinformatic tools to address this challenge (Garber, Grabherr, Guttman, & Trapnell, 2011).

Microarray based analysis in limited numbers of unselected biopsies from patients with COL6-RD has been reported , while RNA-Seq so far has not been performed in biopsies form COL6-RD patients (Paco et al., 2013). To further elucidate pathophysiologic pathways involved at the different histological states of the disease in larger number of biopsies and to compare the two different approaches to expression profiling, we have utilized both microarray and RNA-Seq platforms in the same set of COL6-RD muscle biopsies and normal controls to characterize the progression of COL6-RD as reflected by relative gene expression and transcript patterns. Our study shows that different histological stages of COL6-RD are characterized by the same pattern of gene expression, involving predominantly the upregulation of genes coding for components of the extracellular matrix, which is already evident at a histological stage at which there is no prominent fibrosis evident yet, and the downregulation of muscle specific genes. Our results also suggest that TGFβ1 dependent pathways act as a major drivers of the pathological changes at the skeletal muscle level. Moreover, muscle biopsies from COL6-RD patients are characterized by an increase in alternative spliced isoforms of extracellular matrix-related genes that might contribute to the pathogenesis of the disease.

# 3 - MATERIALS AND METHODS

## 3.1 - Participants

We obtained approval from the National Institute of Neurological Disorders and Stroke (NINDS) Institutional Review Board for this study. Subjects (and/or parents, for minor under the age of 18 years) who agreed to participate provided informed consent. Diagnosis of COL6-RD was made by prior clinical assessments including medical history, neuromuscular examinations, muscle biopsy histological and immunohistochemical studies, and further confirmed by genetic testing (Table 1, at the end of this chapter).

## 3.2 - Histology and images acquisition

Snap frozen muscle biopsies were sectioned with a Thermo Scientific Microm HM 560 cryostat (Thermo Scientific, Wilmington, DE). Ten μm thick cryosections were mounted onto glass slides. Two type of staining were performed: hematoxylin and eosin (H&E), and Gomori trichrome staning (GTC). H&E was performed using a standard protocol. Briefly, sections were incubated for 1 minute in Harris hematoxylin (VWR, Radnor, PA), washed for 2 minutes in running tap water, and incubated in eosin (VWR, Radnor, PA) for 1 minute. The sections were dehydrated in ascending alcohol solutions (50%, 70%, 80%, 95%, 100%) cleared in xylene, and mounted with Permount (, Fisher Scientific, Pittsburgh, PA). GTC was performed using the Gomori Trichrome Kit (Green Collagen) ( Thermo Scientific, Wilmington, DE). Images of specimens were acquired with transmitted-light brightfield using a Zeiss microscope Axio Imager.A2 (Carl Zeiss, Oberkochen, Germany) and AxioCam camera (AxioCam MRc 5; software AxioVision Rel 4.8, Carl Zeiss, Oberkochen, Germany).

## 3.3 - Automated image analysis

To measure the cross sectional area of the muscle fibers, we propose a structured edge detection based hierarchical segmentation algorithm for muscle image segmentation. The structured edge detection, which can better capture inherent muscle image edge structures, is achieved by extending a random decision forest framework. The proposed hierarchical segmentation strategy can segment out muscle cells exhibiting both strong and weak boundaries. An overview of our proposed muscle image segmentation algorithm is shown in Figure 3.

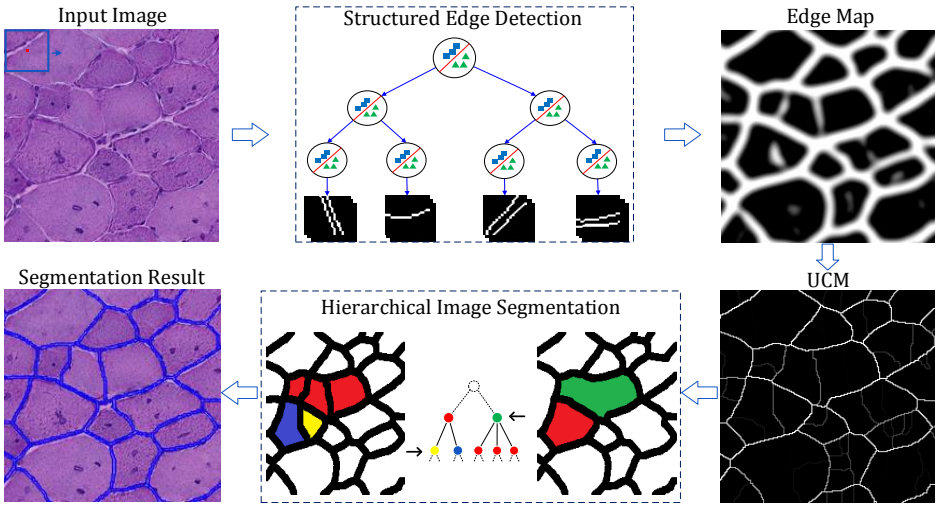



**Figure 3. Overview of the muscle image segmentation algorithm.**
Starting from the captured image, the edge map is built via a structured edge detection algorithm. The Ultrameric Contour Map (UCM) that derives is then subjected to a hierarchical image segmentation algorithm, which results in the final image containing the segmentation results that will be measured to extrapolate the cross sectional area of the muscle fibers

Given an image patch, an edge map is generated by the proposed structured edge detection algorithm; an Ultrametric Contour Map (UCM) is constructed and a set of segmentation candidates is generated by adjusting the thresholds of UCM; and an efficient dynamic programming based subset selection algorithm is then used to choose the best regions for muscle image segmentation based on a constructed tree graph.

The structured edge detection is achieved by extending a random decision forest framework. A decision forest $F = \{T_t\}$ is an ensemble of decision trees $T_t$, which are trained independently on randomly selected samples $S=\{s_i=(x_i, y_i)\}$, where $x_i$ and $y_i$ denote the input features and output labels of the *i-th* training sample, respectively. A decision tree $T_t$ produces the prediction results by recursively branching a feature sample $x$ left or right down the tree until a leaf node is reached. For a decision forest, the predictions $T_t(x)$ from individual trees are combined together using an ensemble model. Majority voting and averaging are common ensemble choices for classification and regression problems, respectively.

During the training of a decision tree, at each node $n$, a split function $h(x)$ is chosen to split the samples Sn into left $S_n^L$ or right $S_n^R$. For multiclass classification, the split function $h(x)$ is optimized by maximizing the information gain.

$$I(S_n) = H(S_n) - (\frac{|S_n^L|}{S_n} H(S_n^L) + \frac{|S_n^R|}{S_n} H(S_n^R))$$

Where $H(S_n)$ is the class entropy function.

The split function $h(x)$ can be an arbitrary classifier. A common choice is a stump function that is found to be computationally efficient and effective in practice. The training procedure continues to split the samples until either a maximum depth is reached, or too few samples are left, or information gain falls below a certain threshold.

Since a decision tree classifier generates the actual prediction at the leaf nodes, more information can be stored at the leaf nodes than class likelihoods. Based on this observation, we propose to store edge structure information at the leaf nodes for structured muscle image edge detection. Different from traditional edge detection algorithms, which take an image patch $x$ as an input and compute the probability of the edge existence at the center pixel $p$, the output of our proposed structured edge detection algorithm is an edge mask around the central pixel $p$ instead of the likelihood value. After learning the decision tree, the median of the edge masks sent to the leaf node will be stored as the leaf node output.

In general, the hierarchical image segmentation consists of two steps: 1) a collection of segmentation candidates is generated by running different segmentation algorithms with varying parameters; 2) Based on some domain-specific criteria, the best subset of the generated segmentation candidates is selected as the final segmentation results. In this work, an Ultrametric Contour Map (UCM) is first constructed on the edge map generated by the proposed structured edge detection algorithm presented in the previous section, then the pool of segmentation candidates is produced by setting different thresholds of UCM. Because of the nice property of UCM where the segmentation results using different thresholds are nested into one another, we can construct a tree graph for this pool of segmentation candidates. The final step is to solve this tree graph based MWIS problem using dynamic programming.

Given a set of segmentation candidates generated with different thresholds using UCM, an undirected and weighted tree graph, $G=(V,E,w)$, is constructed, where $V=\{v_i \mid i=1,2,\ldots,n\}$ represents the nodes with each $v_i$ corresponding to a segmented region $R_i$. The symbol $E$ denotes the edges of the graph. The node weight $w$ is learned via a general random decision forest classifier to represent the likelihood of $R_i$ as a real muscle cell. An adjacent matrix $A=\{a_{ij} \mid i,j=1,\ldots,n\}$ is then built with $a_{ij}=1$ if $R_i$ ($R_j$) is a subset of $R_j$ ($R_i$), and otherwise 0. Denote x as the indicator vector, where its

element is equal to 1 if the corresponding node is selected, otherwise 0. Finally, the constrained subset selection problem is formulated as

$$\mathrm{x}^* = \underset{\mathrm{x}}{\operatorname{argmax}}\, w^T \mathrm{x} \quad \text{such that } \mathrm{x}_i + \mathrm{x}_j \leq 1 \text{ if } a_{ij} = 1$$

Considering the special tree graph structure, above equation can be efficiently solved via the dynamic programming approach with a bottom-to-up strategy.

To quantify the fibrosis from images taken of the Gomori stain we proceeded as follows. Because of staining variations, the intensities of green regions indicating the deposition of collagen in the extracellular matrix might vary considerably across different image regions. In order to precisely measure the green region area, for each image, we manually sample a few points from the green regions, and we call these samples reference points. During testing stage, we compute the intensity differences between each image pixel and its reference points. If the median of these differences was less than a prefixed threshold, this pixel is counted as "green".

After segmentation of green pixels, connected component algorithm is applied to group the green pixels into isolated components. Those components with too few pixels are counted as "noises" and discarded. The green area of the image is computed as the number of all the kept green pixels.

### 3.4 - RNA Isolation and Quality Assessment

Snap frozen muscle biopsies were sectioned with a Thermo Scientific Microm HM 560 cryostat (Thermo Scientific, Wilmington, DE). Fifteen 10 μm cryosections were collected in Trizol (Life Technologies, Carlsbad, CA). The RNA was isolated using standard Trizol protocol, and further purified using miRNeasy Mini Kit (Qiagen, Valencia, CA). After isolation, the RNA concentration was measured using a Nanodrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE). The RNA samples were analyzed using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) to obtain RNA integrity numbers (RIN) as measure of quality (Schroeder et al., 2006). Only samples having a RIN quality ≥ 6 were subject to Microarray analysis RNA-Seq Analysis. See Table 2 for RNA yield and RIN numbers for these samples.

### 3.5 - Microarray Analysis

Labeled cRNA libraries were prepared from isolated total RNA according to Illumina protocols (Illumina, Inc, San Diego, CA). For RNA labeling, 500 ng of total RNA was used in conjunction with the Illumina recommended protocol for hybridization to the HumanHT-12 v4.0 Expression BeadChip. After washing and staining, BeadChips were then scanned using the Illumina BeadArray Reader. Where after, the Illumina GenomeStudio Gene Expression Module was used to generate summarized expression measurements per gene represented on the BeadChip. This expression data was then import into R (http://cran.r-project.org/), Log2 transformed, quantile normalized, then quality inspected via Tukey box plot, covariance-based Principal Component Analysis (PCA) scatter plot, and correlation-based Heat Map. Post quality inspection, the coefficient of variation (CV) and mean expression for each gene was calculated by sample class then modeled using locally weighted scatter plot smoothing (CV ~ mean). Resulting model fits were then inspected and used to define "system noise" as the mean expression value across sample classes at which the linear relationship between CV (i.e., noise) and mean expression (i.e., signal) is grossly lost. Genes not having at least one normalized value (Quantile(Log2(RPKM+2))) greater than "system noise" were discarded as noise-biased. While, normalized values for remaining genes were floored to equal "system noise" if less than "system noise". To identify differences in normalized values (i.e., gene expression) between CMD samples and Control samples, the Welch-modified t-test was applied under Benjamin–Hochberg (BH) False-Discovery Rate (FDR) Multiple Comparison Correction (MCC) condition. Such that, genes with a corrected $P < 0.05$ by this test and an absolute fold-difference $\geq 1.5X$ between class means were those considered to be differentially expressed between CMD and Control. Annotations for these genes were obtained from IPA (http://www.ingenuity.com/). Where after, IPA was used again to determine what biological functions and pathways are enriched for these genes; along with what networks of genes they have the most connectivity with. Comparison and intersection of results with those obtained by RNA-Seq Analysis was accomplished using the IPA-assigned gene symbol.

### 3.6 - RNA-Seq Analysis

cDNA libraries were prepared from isolated RNA according to Illumina protocols (Illumina, Inc, San Diego, CA) and sequenced by Beckman Coulter Genomics (Danvers, Massachusetts). Briefly, mRNA libraries were constructed from 150 ng of total RNA using the Illumina TruSeq RNA Sample Prep Kits, version 2. The resulting cDNA was fragmented using a Covaris E210. Library amplification was performed using 12-15 cycles to minimize the risk of over-amplification.

Unique barcode adapters were applied to each library. Equal volumes of individual libraries were pool and run on a MiSeq. The libraries were then repooled based on the MiSeq demultiplexing results and sequenced on a HiSeq 2000 with ver 3 flow cells and sequencing reagents. On average, 69 million 100-base read pairs were generated for each individual library. Data was processed using RTA 1.13.48 and deplexed via CASAVA 1.8.2. Post deplexing, resulting .fastq files were trimmed to remove 3' occurring adaptor sequences using the FASTQ/A Clipper tool (http://hannonlab.cshl.edu). Adaptor trimmed .fastq files were then import into the CLCbio Genomics Workbench (www.clcbio.com) for quality inspection, quality trimming, filtering, and mapping. Per quality trimming, 15bp from the 5' end of each read for each sample was globally removed along with 1bp from the 3' end. Bases with a call accuracy < 95% were also removed. Per filtering, read pairs having more than two ambiguities in at least one read were discarded along with pairs < 15bp in length. Per mapping, the "RNA-Seq" tool was used under default parameters to align reads in pairs to the human genome (hg19). Post mapping, the number of aligned reads falling in each annotated "gene" were enumerated then converted into number of "Reads Per Kilo base of transcript Million mapped reads" ("RPKM"). RPKM values were then import into R (http://www.r-project.org/), pedestalled by 2 (RPKM+2), log (base 2) transformed (Log2(RPKM+2), and filtered to remove genes not having a transformed value >1 for at least one sample. Post filtering, transformed values for remaining genes were quantile normalized (Quantile(Log2(RPKM+2))) then quality inspected via Tukey box plot, covariance-based principal component analysis (PCA) scatter plot, and correlation-based Heat map. Post quality inspection, the coefficient of variation (CV) and mean expression for each gene was calculated by sample class then modeled via locally weighted scatter plot smoothing (CV ~ mean). Resulting model fits were then inspected and used to define "system noise" as the mean expression value across sample classes at which the linear relationship between CV (i.e., noise) and mean expression (i.e., signal) is grossly lost. Genes not having at least one normalized value (Quantile(Log2(RPKM+2))) greater than "system noise" were discarded as noise-biased. While, normalized values for remaining genes were floored to equal "system noise" if less than "system noise". To identify differences in normalized values (i.e., gene expression) between CMD and Control, the Welch-modified t-test was applied under Benjamin–Hochberg (BH) False-Discovery Rate (FDR) Multiple Comparison Correction (MCC) condition. Such that, genes with a corrected $P < 0.05$ by this test and an absolute fold-difference >= 1.5X between class means were those considered to be differentially expressed between CMD and Control. Annotations for these genes were obtained from IPA (http://www.ingenuity.com/). Where after, IPA was used again to determine what biological functions and pathways are enriched for these genes; along with what

networks of genes they have the most connectivity with. Comparison and intersection of results with those obtained by Microarray Analysis was accomplished using the IPA-assigned gene symbol. Finally, the same methods described were repeated using "transcript-level" enumerations in place of "gene-level" enumerations; providing for a list of differentially expressed transcripts occurring between CMD and control.

### 3.7 - qRT-PCR expression validation

We synthesized cDNA from 300 ng of total RNA using the pScript XLT cDNA SuperMix (Quanta Biosciences, Gaithersburg, MD). We measured cDNA abundance using Sybr Green on an Applied Biosystems QuantStudio 6 qRT-PCR system (Life Technologies, Carlsbad, CA). We performed serial dilutions of cDNA to find primer pairs with 100% efficiency and a single product on the dissociation curve. We used TBP as a normalization gene. Primer pairs are listed in Table 3. As biological replicates, we used 6 COL6-RD samples and 5 control samples. All the experiments were conducted in triplicates. Fold-change was assessed determining the ΔΔCt between the normalization gene and the target gene.

### 3.8 - Immunoblotting

Snap frozen muscle biopsies were sectioned with a Thermo Scientific Microm HM 560 cryostat (Thermo Scientific, Wilmington, DE). Twenty-to-thirty 10 μm cryosections were collected and mechanically disrupted and lysed in 100 μl of NP40 lysis buffer (1M Tris-HCl pH 7.4, 1.5M NaCl, 0.5M EDTA pH 7.4, 1% NP40) with protease inhibitors (chymostatin, leupeptin, pepstatin A, antipain) and phosphatase inhibitors (Phostop, Roche, Mannheim, Germany). The cell extracts were centrifuged for 15 minutes at 14,000 g at 4 ºC. Protein concentration was assessed using BCA protein assay kit (Millipore, Billerica, MA). Thirty μg of total protein was separated in NuPage 4-12% Bis-Tris gel (Life Technologies, Carlsbad, CA) and transferred on PVDF membrane for probing. The primary antibodies were used at the dilution of 1:1000. The primary antibodies used were: THBS4 (AF2390, R&D Systems, Minneapolis, MN), BECN1 (3738, Cell Signaling, Danvers, MA), LC3B (3868, Cell Signaling, Danvers, MA), DCN (AF143, R&D Systems, Minneapolis, MN), BGN (LF-51, kindly provided by Dr. Larry Fisher, NIDCR/NIH), Phospho-Smad2 (3101, Cell Signaling, Danvers, MA), Smad2 (5339, Cell Signaling, Danvers, MA ). Bound antibodies were detected with the appropriate secondary antibodies HRP-conjugated. The results were normalized to GAPDH (MAB374, Millipore, Billerica, MA).

| Patients | | Mutation | | | | |
|---|---|---|---|---|---|---|
| **ID** | **Sex** | **Gene** | **Exon** | **cDNA** | **Protein** | **Effect** |
| UC179 | M | COL6A1 | 14 | c.1057+1G>A | Intronic | Splice Site Mutation |
| UC21 | M | COL6A3 | 16 | c.6610+2T>del | Intronic | Splice Site Mutation |
| UC397.1 | F | COL6A1 | 1 | c.46T>C | p.Trp16Arg | Uncertain |
| | | COL6A1 | 19 | c.1290_1291delAG | p.Gly431ThrfsX16 | Frame shift |
| UC264 | M | COL6A2 | 6 | c.813G>A | p.Gly271Asp | Triple Helical Domain |
| UC220 | M | COL6A2 | 24 | c.1771-25A>G | Intronic | Splice Site Mutation |
| | | COL6A2 | 26 | c.1970-9G>A | Intronic | Splice Site Mutation |
| UC51 | M | COL6A2 | 28 | c.2793_2821del29 | p.Ser931_940del | Frame shift |
| UC222 | F | COL6A2 | 6 | C.856-3C>G | Intronic | Splice Site Mutation |
| | | COL6A2 | 26 | c.2386A>T | p.Lys796Stop | Premature Stop Codon |
| UC13 | M | COL6A2 | 26 | c.2170C>T | p.Arg724Cys | Uncertain |
| | | COL6A2 | 26 | c.2411T>G | p.Val804Gly | Uncertain |
| | | COL6A3 | 15 | c.6156+1G>A | Intronic | Splice Site Mutation |
| UC30 | M | COL6A1 | 13 | c.1002+1G>A | Intronic | Splice Site Mutation |
| UC99 | F | COL6A1 | 11 | c.904-1G>C | Intronic | Splice Site Mutation |
| UC95 | F | COL6A3 | 17 | c.6229G>C | p.Gly2077Arg | Triple Helical Domain |
| UC338 | M | COL6A2 | 26 | c.2197G>A | p.Gly733Arg | Triple Helical Domain |
| | | COL6A2 | 28 | c.2917G>A | p.Val973Met | Uncertain |
| UC100 | M | COL6A3 | 16 | c.6610+1G>A | Intronic | Splice Site Mutation |
| UC54 | F | COL6A3 | 16 | c.6610+1G>A | Intronic | Splice Site Mutation |
| UC56 | F | COL6A3 | 18 | c.6408+9C>G | p.Arg2136_Gly2137ins3 | Splice Site Mutation |
| UC384 | M | COL6A1 | 10 | c.877G>A | p.Gly293Arg | Triple Helical Domain |
| | | COL6A1 | 10 | c.859-10A>G | Intronic | Uncertain |
| UC43 | M | COL6A2 | 26 | c.1970-3C>A | Intronic | Splice Site Mutation |
| | | COL6A2 | 9 | c.954+20_+55del36 | Uncertain | Uncertain |
| UC98 | M | COL6A1 | 21 | c.1451C>del | Uncertain | Frame shift |
| UC50 | M | COL6A3 | 15 | c.6156+2T>C | Intronic | Splice Site Mutation |
| UC129 | M | COL6A3 | 15 | c6156+1G>A | Intronic | Splice Site Mutation |
| UC77 | M | COL6A1 | 12 | c.957+1G>A | Intronic | Splice Site Mutation |
| UC357 | F | COL6A2 | 10 | c.954_955ins11 | Intronic | Splice Site Mutation |

**Table 1. COL6RD sample identification code and corresponding cDNA and protein pathogenic mutations.**

| Sample ID | Concentration (ng/ µl) | RIN | Sample ID | Concentration (ng/ µl) | RIN |
|---|---|---|---|---|---|
| M13-1 | 140.7 | 8.2 | UC220 | 204.0 | 8.5 |
| M13-7 | 113.5 | 8.4 | UC51 | 63.6 | 8.3 |
| M13-16 | 117.7 | 7.3 | UC222 | 65.0 | 8.3 |
| M13-18 | 83.5 | 7.2 | UC13 | 124.6 | 8.2 |
| M13-23 | 68.9 | 7.1 | UC30 | 86.6 | 7.9 |
| M13-25 | 60.5 | 7.1 | UC99 | 162.5 | 8.7 |
| M13-30 | 93.8 | 8.3 | UC95 | 97.9 | 7.4 |
| M13-33 | 156.5 | 8.2 | UC338 | 62.8 | 7.3 |
| M13-35 | 87.8 | 7.7 | UC100 | 184.9 | 8.9 |
| M13-40 | 75.8 | 7.1 | UC54 | 95.5 | 7.9 |
| M13-43 | 133.5 | 8.1 | UC56 | 104.3 | 7.5 |
| M13-45 | 67.3 | 7.6 | UC384 | 108.8 | 8.4 |
| M13-47 | 112.9 | 7.3 | UC43 | 126.1 | 7.10 |
| M13-49 | 140.6 | 8 | UC98 | 61.4 | 7.8 |
| UC179 | 104.9 | 8 | UC50 | 92.3 | 7.7 |
| UC21 | 139.0 | 9 | UC129 | 141.9 | 8.3 |
| UC397.1 | 60.2 | 7.8 | UC77 | 54.0 | 7.8 |
| UC264 | 91.5 | 8.4 | UC357 | 81.4 | 7.7 |

**Table 2. Controls (M) and COL6-RD (UC) samples with respective RNA concentration (ng/µl) and quality assessment based on the RNA Integrity Number (RIN).**

| Gene Symbol | Forward Primer | Reverse Primer |
|---|---|---|
| THBS4 | CTTGCGGTCCTCTCAAGTTT | CGGGTAGCAGGGATGGTATT |
| TGFB1 | TGGAAACCCACAACGAAATCT | GTATCGCCAGGAATTGTTGCT |
| COL1A2 | TGAAGATGGTCACCCTGGAA | CCCCTAATGCCTTTGAAGCC |
| COL3A1 | TCAAGGGTGAAAGTGGGAAAC | CACGATCACCCTTGCCAC |
| CTGF | GGAAGAGAACATTAAGAAGGGCA | CCGTCGGTACATACTCCACA |
| BGN | CAGCACCTCTACGCCCTC | GCAGTTCATGTTCCGGAGC |
| FN1 | AGTCTTTACCACACTGCAGC | ACAACGATGCTTCCTGAGTC |
| ATROGIN1 | ACCAAAACTCAGTATTTCCACCA | AAAGGCAGGTCAGTGAAGGT |
| BECLIN1 | CAGGAGGAAGCTCAGTATCAGA | TTGTGCCAAACTGTCCACTG |
| BNIP3 | TCTTTAAACACCCGAAGCGC | CGCCTTCCAATATAGATCCCC |
| TP53 | GTTTCCGTCTGGGCTTCT | ACCTCAGGCGGCTCATAG |
| CDK1N | CAGACCAGCATGACAGATTTCT | CGGCCAGGGTATGTACATGA |
| CILP | ACCCACATCTACCTCTTGCC | CACTCACCAGGGCTCTCC |
| MGP | AGCCTGATCCTTCTTGCCAT | CGTTCTCGGATCCTCTCTTG |
| LUM | CCTTCTGGTCTCCCTGTCTC | GATCCAGCTCAACCAGGGAT |
| MYH1 | CGTCAGTCAAGGGGAGATCA | TTGTCAGCAACTTCAGTGCC |
| TBP | CCGAATATAATCCCAAGCGGTT | GCTCCTGTGCACACCATTTT |

**Table 3. Primers used for qRT-PCR analyses.**

# 4 - RESULTS

## 4.1 - Morphological characterization

In order to assign histological stages of severity to the biopsies included in the molecular analysis we first performed a basic histological review of the samples.

Twenty two snap-frozen COL6RD skeletal muscle biopsies were analyzed using basic histology staining: hematoxylin and eosin (H&E) and Gömöri trichrome staining (GTC). Fifteen patients were males, seven were female. All of the patients whose muscle biopsies were studied have a genetically confirmed diagnosis of collagen VI-related dystrophy with pathogenic mutations occurring in one of the three genes coding for collagen type VI (COL6A1, COL6A2, COL6A3) (Table 1). Fifteen patients carried de novo dominant mutations, while 7 of the patients have recessively inherited mutations. The mutations were distributed as follow: 31% on COL6A1, 44% on COL6A2, 25% on COL6A3. We arbitrarily formed three groups based on the severity of the histological changes. The first group, COL6RD-1, was characterized by very mild histological changes, mainly represented by muscle fiber atrophy; the second group, COL6RD-2, was defined by muscle fiber atrophy and moderate fibrosis. Fibrosis was defined moderate when 40%-60% of the individual muscle fibers were surrounded by a rim of connective tissue identified by the color blue in the GTC stain, with overall preservation of the skeletal muscle architecture. The third group, COL6RD-3, was defined by muscle fiber atrophy, severe fibrosis, and fat infiltration. Fibrosis was considered severe when more than 60% of the individual muscle fibers were surrounded by connective tissue, together with skeletal muscle architecture disruption. We visually assigned the muscle biopsies to any of these groups and we used the automated image analysis algorithm described in the methods to quantify cross sectional area and fibrosis. The reduction in the cross sectional area of the muscle fibers was evident and statistically significant in all of the three groups compared to normal controls. We also noted a change in the fiber size distribution between controls and COL6RD muscle biopsies. The patients muscle samples had a prevalence of muscle fibers with cross sectional area smaller than 1000 $\mu m^2$, while the normal controls had a prevalence of muscle fibers with cross sectional area between 1000 $\mu m^2$ and 2000 $\mu m^2$. The quantification of connective tissue also followed exactly the way we assigned the patients muscle biopsies to the three groups. As expected, COL6RD muscle biopsies showed higher amount of connective tissue when compared to normal controls, increasing proceeding from COL6RD-1, COL6RD-2, to COL6RD-3. Results are shown in Figure 4.

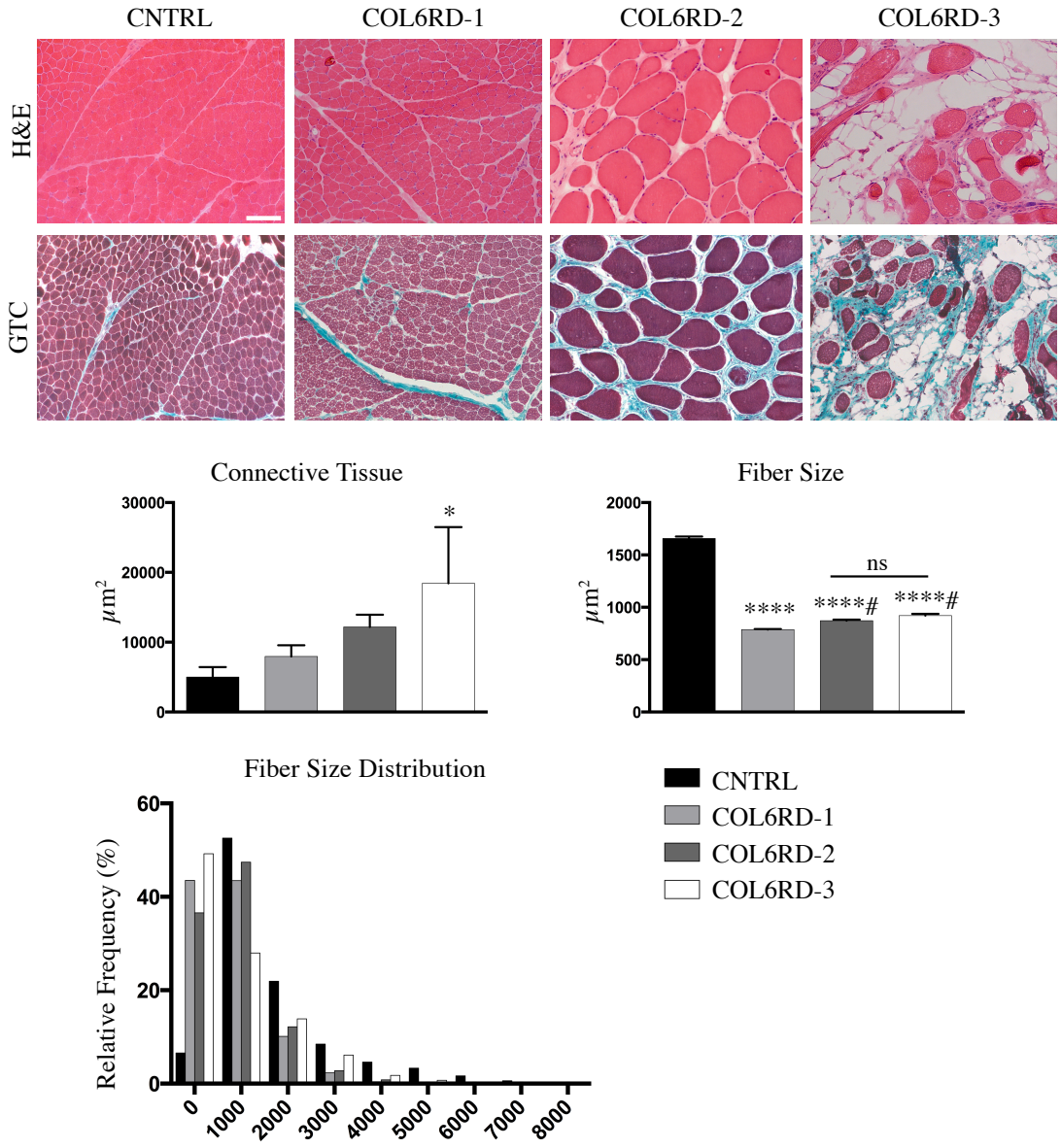



**Figure 4. Representative histochemical stains of frozen muscle sections.**

We used Hematoxylin & Eosin and Gomori Trichrome staining to visualize basic histology features such as muscle fiber atrophy, and connective tissue deposition, that allowed us to group the patient samples into three groups according to the severity of the muscle histology: COL6RD-1,2,3. Quantification of cross sectional area/ fiber size, amount of connective tissue, and distribution of fiber size according to the grouping system is shown on the lower panel. The values are expressed as mean ± S.E.M. The statistical significance was established using the non-parametric Mann-Whitney test. * indicates p-value < 0.05; *** indicates p-values < 0.0001 when compared to the control group. # indicates p-value < 0.0001 when compared to COL6RD-1 group. Scale bar corresponds to 100 μm

### 4.2 - Differential gene expression analysis

We next asked whether the relative severity of histological changes in the three groups could be correlated to specific patterns of differential gene expression in order to investigate whether there were different molecular signatures evident at different stages of the disease process. We used microarray-based gene expression profiling and RNA-Seq to compare the transcriptome of the 22 COL6RD muscle biopsies against 15 aged-matched controls.

Row gene expression data from microarray analysis were normalized and low intensity values below the noise threshold were filtered out. After filtering, microarray analysis resulted in 11,172 gene probes present at least in one sample with an expression value greater than the noise threshold, corresponding to a total of 8234 human annotated genes. Of these, 648 were differentially regulated between normal controls and COL6RD muscle biopsies, with absolute expression level of 1.5-fold or greater and p-value $< 0.05$. Of these, 479 differentially regulated genes were upregulated, while 169 were downregulated.

RNA-Seq generated an average of 50 million reads per samples. Quality inspection of the reads showed an expected per-sample GC content peak around 60%, ambiguity base content at nearly 0%, and average PHRED score near 40 across all the samples, which is indicative of the high quality of the sequencing. Filtering and trimming of the reads were uniform throughout the samples and, most importantly, preserved 70-80% of the total reads sequenced. Moreover, 99% of the post-trimmed and filtered reads were successfully mapped to the human genome. After trimming and filtering, RNA-Seq identified 2889 human annotated genes. Statistical significance testing determined that a total of 248 genes were differentially expressed between control and COL6RD muscle biopsies, with absolute expression level of 1.5-fold or greater and p-value $< 0.05$. Of these, 170 differentially regulated genes were upregulated, while 78 were downregulated.

Only 138 differentially regulated genes were commonly detected by microarray and RNA-Seq. For these, there was good concordance in direction ($R^2=1$) and magnitude ($R^2=0.84$) of the fold change (Figure 5c).

Hierarchical heat maps and Principal Component Analysis (PCA) were obtained for both microarray and RNA-Seq. Analysis of the hierarchical heat map and the PCA allowed the identification of unique gene expression patterns clearly differentiating controls and COL6RD muscle biopsies (Figure 5a, b). However, no clear distinction was evident within the COL6RD muscle biopsies correlating with the morphological groups the muscle biopsies had previously been assigned to.

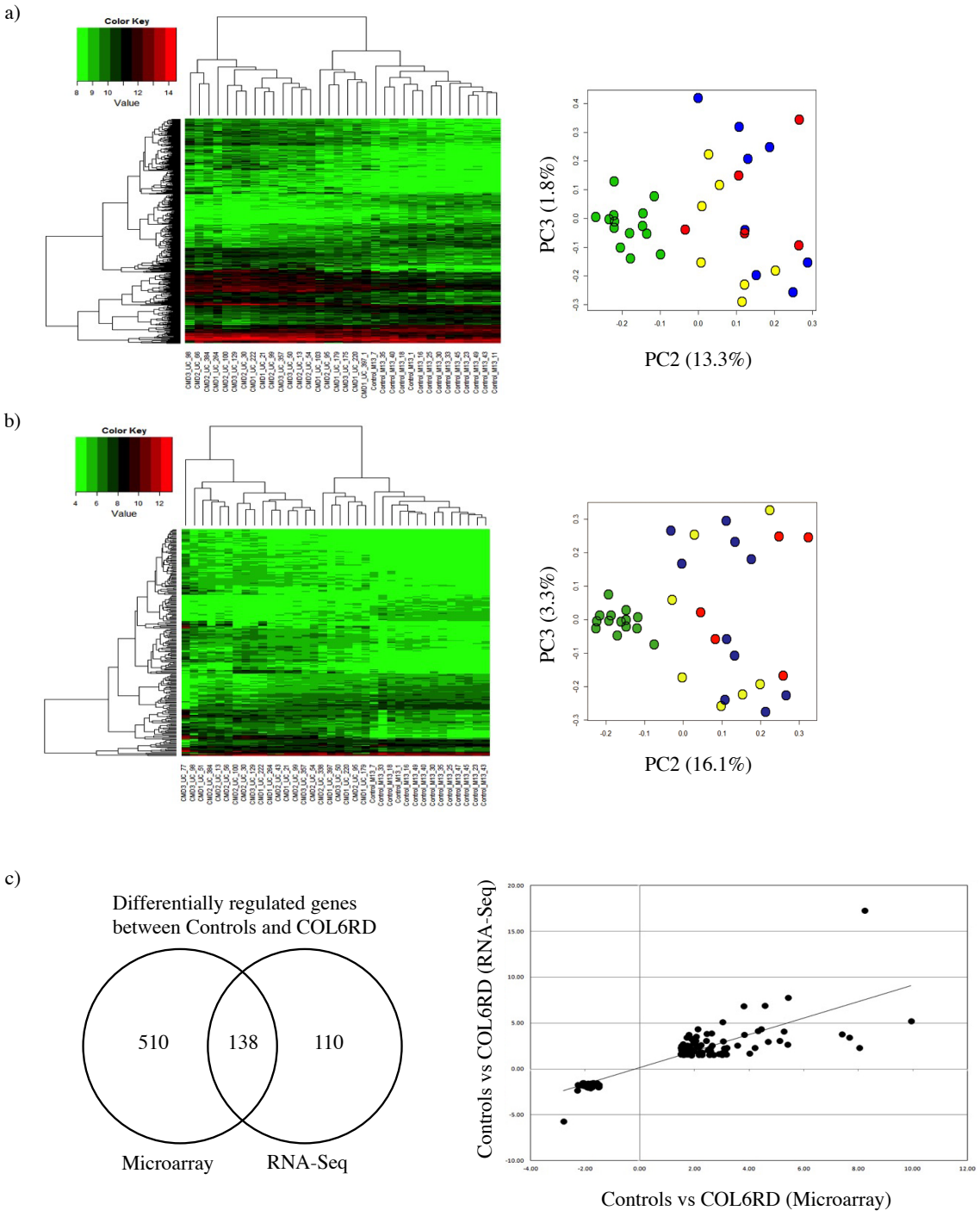



**Figure 5. Summary of the microarray and RNA-seq findings.**

Panel a and b show post significance analysis hierarchical heat map obtain from microarray (a) and RNA-Seq (b) gene expression profiling. Principal component analysis of the gene expression levels obtain by microarray and RNA-Seq is also shown. The two PCA analyses show clearly that the controls cluster together and segregate from the COL6RD samples. However, there is no clear clustering of the COL6RD samples by histology groups, as they are all scattered and mixed. Panel c shows Venn diagram showing the total number of differentially regulated genes (p-value <0.05; Fold change magnitude = 1.5) obtained from either microarray or RNA-Seq platforms. In the central part are the dysregulated genes identifies by both techniques. On the right, comparison of direction and fold change of the differentially regulated genes commonly detected by microarray and RNA-Seq. The plot shows a high concordance of direction ($R^2$=1) and magnitude ($R^2$=0.84)

We next more closely analyzed all the genes that were differentially regulated in either microarray or RNA-Seq gene expression analyses. We combined the list of differentially regulated genes produced by microarray and RNA-seq in a global gene list of uniquely occurring genes. For each gene we considered only the highest value of expression. This process generated 758 differentially regulated genes which were submitted to the web-based Database for Annotation, Visualization and Integrated Discovery (DAVID) for gene ontology analysis. Ranked by p-value and gene count, the upregulated genes were most commonly represented by the gene ontology terms extracellular matrix (p-value=6.18E-38, gene count=69), proteinaceous extracellular matrix (p-value=4.46E-36, gene count= 65), extracellular region part (p-value=1.66E-30, gene count=100), and extracellular region (p-value=7.68E-23, gene count=7.68). In contrast, the downregulated genes were represented by the gene ontology terms mitochondrion (p-value=3.11E-14, gene count=53), generation of precursor metabolites and energy (p-value=2.16E-12, gene count=27), glucose metabolic process (p-value=1.91E-10, gene count=18), and myofibril (p-value=1.25E-09, gene count=15) (Table 4).

| **Upregulated Genes** | | | **Downregulated Genes** | | |
|---|---|---|---|---|---|
| **Gene Ontology Term** | **Gene Count** | **p-value** | **Gene Ontology Term** | **Gene Count** | **p-value** |
| Extracellular matrix | 69 | 6.18E-38 | Mitochondrion | 53 | 3.11E-14 |
| Proteinaceous extracellular matrix | 65 | 4.46E-36 | Generation of precursor metabolites and energy | 27 | 2.16E-12 |
| Extracellular region part | 100 | 1.66E-30 | Glucose metabolic process | 18 | 1.91E-10 |
| Extracellular region | 135 | 7.68E-23 | Myofibril | 15 | 1.25E-09 |
| Extracellular matrix part | 31 | 1.46E-20 | Mitochondrial part | 32 | 1.50E-09 |
| Cell adhesion | 65 | 1.04E-18 | Contractile fiber part | 15 | 1.58E-09 |
| Biological adhesion | 65 | 1.12E-18 | Sarcomere | 14 | 2.64E-09 |
| Extracellular matrix organization | 27 | 1.36E-18 | Contractile fiber | 15 | 3.93E-09 |
| Glycosaminoglycan binding | 29 | 4.10E-17 | Hexose metabolic process | 18 | 6.51E-09 |
| Extracellular matrix structural constituent | 23 | 4.25E-16 | Monosaccharide metabolic process | 19 | 9.28E-09 |
| Pattern binding | 29 | 6.62E-16 | Muscle contraction | 16 | 1.31E-08 |
| Polysaccharide binding | 29 | 6.62E-16 | Energy derivation by oxidation of organic compounds | 15 | 4.53E-08 |
| Structural molecule activity | 57 | 1.46E-15 | Muscle system process | 16 | 4.66E-08 |
| Extracellular structure organization | 29 | 2.50E-15 | Striated muscle contraction | 9 | 4.31E-07 |
| Growth factor binding | 23 | 4.14E-14 | Glycogen metabolic process | 8 | 8.57E-07 |

**Table 4. Gene ontology analysis.**
Gene ontology analysis using DAVID (Huang da, Sherman, & Lempicki, 2009a, 2009b) performed on the list of differentially regulated genes obtained from the combination of the uniquely occurring genes from microarray and RNA-Seq data. Of the 758 differentially regulated gene, 711 were accepted for gene ontology analysis. Here, we show the top 30 gene ontology categories, classified by p-value. The most represented term for the upregulated gene is related to the extracellular matrix biogenesis and remodeling. The downregulated genes predominantly relate to skeletal muscle structure and function (i.e. muscle organ development, contractile fiber part), energy production, and metabolism.

We next evaluated the gene expression differences between controls and COL6RD muscle biopsies considering predicted cellular localization and molecular function. Again, in COL6RD muscle biopsies we found that a disproportionate number of the upregulated gene products is predicted to be localized in the extracellular matrix (117 upregulated genes vs 5 downregulated genes), plasma membrane (86 upregulated genes vs 33 downregulated genes) and nucleus (71 upregulated genes vs 35 downregulated genes) (Figure 6 a).

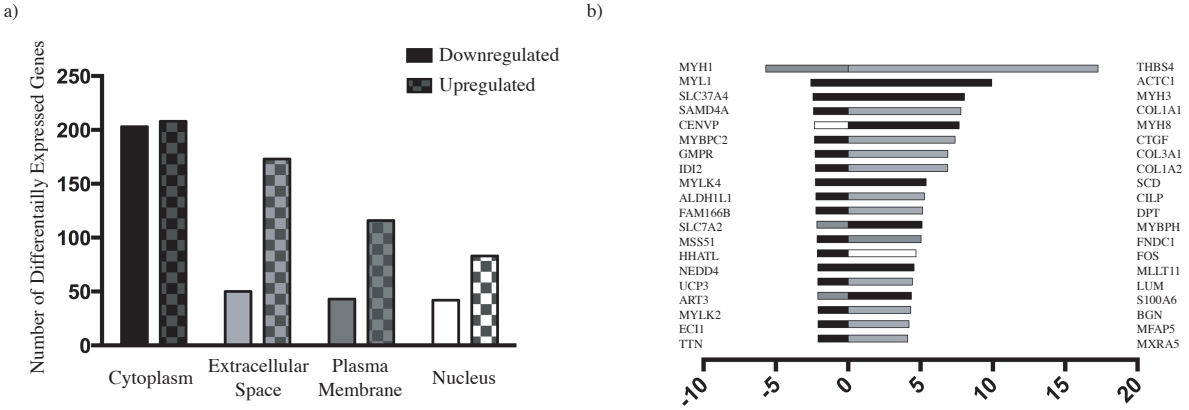



**Figure 6. Cellular Distribution of the differentially regulated genes obtained by the combination of microarray and RNA-Seq data.**

Panel a shows cellular localization of up- and downregulated genes in COL6RD vs controls. The plot shows how there are more upregulated than downregulated genes. The majority of the upregulated genes are related to the extracellular matrix and the plasma membrane. Panel b shows the top 20 changing genes listed according to fold-change and color-coded according to their cellular localization.

Using the same combined gene list (microarray and RNAseq-based) sorted by magnitude of the gene expression level we identified thrombospondin 4 (THBS4) as the single gene with the greatest difference between COL6RD and control muscle biopsies (fold-change = 17.28; corrected p-value = 2.59E-09). Among the 20 highest upregulated genes there were also 4 genes annotated to be involved in cartilage and bone formation, namely CILP1, MGP, LUM, and ASPN. Among the twenty most upregulated genes were also find CTGF, COL3A1, and COL1A2, which are associated with connective tissue deposition and, ultimately, with fibrosis (Figure 6 b).

Using Ingenuity Pathway Analysis, we searched for cellular function annotations enriched in the dataset, considering both upregulated and downregulated genes. We considered only annotations with z-score >2 or <-2, and p-value <0.05. Based on z-score and p-value, the cellular function annotations predicted to be increased in COL6RD muscle biopsies were cell movement (z-score=3.342, p-value=1.64E-20), migration of cells (z-score=3.135p-value=7.23E-20), leukocyte migration (z-score=3.093, p-value=1.12E-06). The cellular function annotations predicted to be

decreased are concentration of ATP (z-score=-2.178, p-value=1.97E-04) and colony formation of tumor cell lines (z-score=-2.470, p-value=1.23E-06) (Table 5).

| Cellular Function Annotation | p-Value | Predicted Activation State | Activation z-score | Cellular Function Annotation | p-Value | Predicted Activation State | Activation z-score |
|---|---|---|---|---|---|---|---|
| cell movement | 1.64E-20 | Increased | 3.342 | cell movement of embryonic cell lines | 1.02E-04 | Increased | 2.508 |
| migration of cells | 7.23E-20 | Increased | 3.135 | cell movement of monocytes | 2.15E-07 | Increased | 2.496 |
| leukocyte migration | 1.12E-06 | Increased | 3.093 | adhesion of vascular endothelial cells | 2.03E-04 | Increased | 2.493 |
| differentiation of cells | 4.78E-10 | Increased | 2.981 | differentiation of tumor cell lines | 8.60E-06 | Increased | 2.465 |
| adhesion of kidney cells | 1.26E-05 | Increased | 2.949 | migration of breast cancer cell lines | 6.08E-08 | Increased | 2.429 |
| cell movement of phagocytes | 8.40E-07 | Increased | 2.902 | invasion of tumor cell lines | 2.63E-16 | Increased | 2.353 |
| cell movement of tumor cell lines | 2.93E-15 | Increased | 2.880 | adhesion of connective tissue cells | 7.11E-07 | Increased | 2.350 |
| chemotaxis of cells | 3.13E-06 | Increased | 2.872 | cell movement of mononuclear leukocytes | 4.83E-07 | Increased | 2.333 |
| migration of tumor cell lines | 3.91E-14 | Increased | 2.796 | organization of actin cytoskeleton | 1.96E-05 | Increased | 2.236 |
| adhesion of embryonic cells | 6.27E-06 | Increased | 2.756 | cell movement of myeloid cells | 3.55E-05 | Increased | 2.231 |
| adhesion of embryonic cell lines | 2.64E-05 | Increased | 2.756 | cell survival | 1.23E-05 | Increased | 2.205 |
| cell movement of leukocytes | 1.15E-06 | Increased | 2.723 | cell spreading of epithelial cell lines | 3.10E-05 | Increased | 2.190 |
| organismal death | 5.90E-09 | Increased | 2.646 | binding of tumor cell lines | 3.59E-13 | Increased | 2.180 |
| invasion of cells | 8.66E-20 | Increased | 2.644 | cell spreading | 4.99E-08 | Increased | 2.141 |
| adhesion of epithelial cell lines | 2.56E-04 | Increased | 2.599 | binding of leukemia cell lines | 1.38E-09 | Increased | 2.113 |
| adhesion of kidney cell lines | 1.15E-04 | Increased | 2.599 | migration of monocytes | 4.59E-07 | Increased | 2.095 |
| migration of phagocytes | 1.87E-07 | Increased | 2.591 | binding of leukocytes | 1.53E-04 | Increased | 2.029 |
| cell movement of pancreatic cancer cell lines | 1.15E-04 | Increased | 2.587 | concentration of ATP | 1.97E-04 | Decreased | -2.178 |
| cell movement of breast cancer cell lines | 1.12E-08 | Increased | 2.532 | colony formation of tumor cell lines | 1.23E-06 | Decreased | -2.470 |

**Table 5. Cellular function enrichment.**

Cellular functions enrichment was performed using Ingenuity Pathway Analysis. The cellular function annotations are presented classified by p-value and activation z-score.

### 4.3 - Differentially expressed splice isoforms analysis

We used the list of differentially regulated genes generated uniquely from the RNA-Seq to determine differences in the expression of well annotated splice isoforms (i.e. differences in transcript originating from the same gene). We detected more than one differentially regulated splice isoform for 43 of the 248 differentially expressed genes. The number of different splice isoforms per gene was highly variable and ranging from 2-to-12 transcripts per gene, with DECORIN showing 12 differentially regulated splice isoforms. We submitted the transcripts to Ensemble Database for better characterization. We found that about 61% of the splice isoforms were protein coding (77/126), while 39% of the detected differentially expressed splice isoform was non protein coding. This non coding portion was further subdivided in isoforms that were not coding for protein because of retained introns (22%), and nonsense mediated decay (8.7%). However, the remaining 7.1% of non coding transcripts was still processed by the cell, even if they would not result in any protein product (Figure 7).

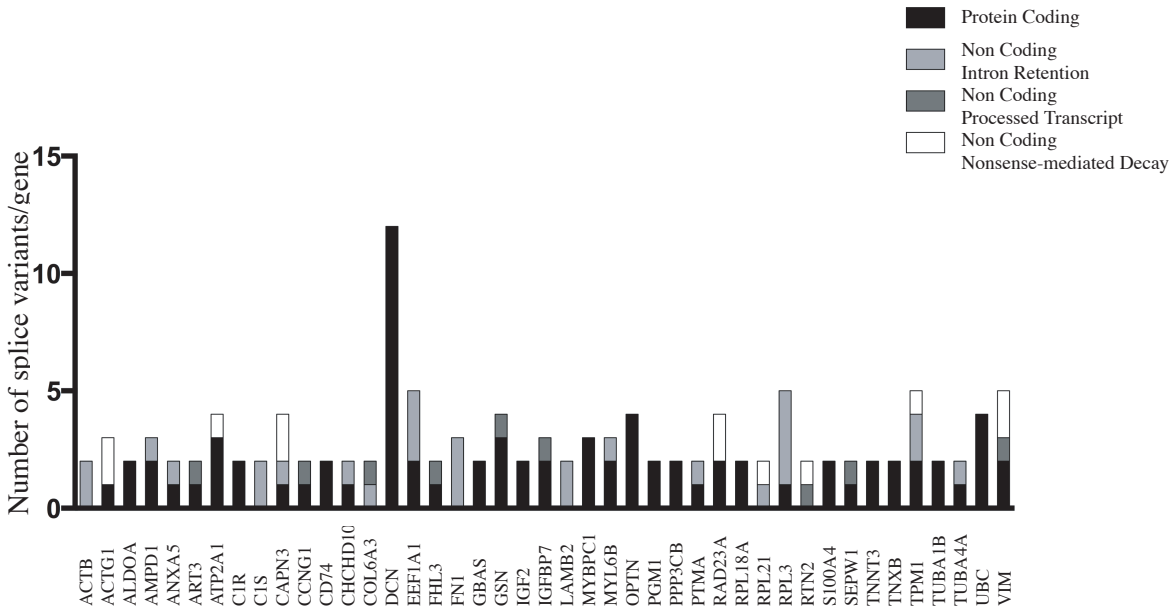



**Figure 7. Splice variants differentially expressed in COL6RD.**

We show the differentially regulated genes that have more than one splice variant significantly differentially regulated in COL6RD compared to normal controls, and the number of transcript per gene. Each bar is proportionally color-coded according to the predicted function of the splice variants represented.

We next performed gene ontology analysis on the 43 represented genes with differential splice isoform expression. The upregulated splice isoforms were enriched in terms related to extracellular matrix biogenesis, and structural molecule activity. The downregulated splice isoforms were related to skeletal muscle function and structure (Table 6).

| Upregulated | | | Downregulated | | |
|---|---|---|---|---|---|
| **Gene Ontology Term** | **Gene Count** | **p-value** | **Gene Ontology Term** | **Gene Count** | **p-value** |
| Structural molecule activity | 11 | 2.73E-08 | Myofibril | 7 | 2.79E-09 |
| Response to wounding | 7 | 2.01E-04 | Contractile fiber | 7 | 4.71E-09 |
| Structural constituent of cytoskeleton | 4 | 2.53E-04 | Sarcomere | 6 | 9.77E-08 |
| Focal adhesion | 6 | 2.81E-04 | Contractile fiber part | 6 | 2.00E-07 |
| Cellular component morphogenesis | 6 | 4.61E-04 | I band | 4 | 4.06E-05 |
| Acute inflammatory response | 4 | 5.88E-04 | Actin cytoskeleton | 5 | 2.86E-04 |
| Extracellular region part | 8 | 6.01E-04 | Regulation of system process | 5 | 6.34E-04 |
| Translational elongation | 4 | 6.42E-04 | Actin binding | 5 | 7.17E-04 |
| Cell morphogenesis involved in differentiation | 5 | 6.98E-04 | Muscle contraction | 4 | 1.02E-03 |
| Extracellular matrix part | 4 | 8.81E-04 | Structural molecule activity | 6 | 1.04E-03 |
| Proteinaceous extracellular matrix | 5 | 1.65E-03 | Muscle system process | 4 | 1.34E-03 |
| ECM-receptor interaction | 4 | 2.09E-03 | Striated muscle contraction | 3 | 1.67E-03 |
| Extracellular matrix | 5 | 2.17E-03 | Striated muscle cell development | 3 | 2.13E-03 |
| Extracellular region | 10 | 2.80E-03 | Muscle cell development | 3 | 2.47E-03 |
| Cell morphogenesis | 5 | 2.81E-03 | Muscle organ development | 4 | 2.57E-03 |

**Table 6. Gene ontology analysis on the splice variants.**
Gene ontology analysis using DAVID performed on the genes having >2 differentially regulated splice variants in COL6RD vs controls. Of the 46 genes submitted to DAVID, 43 were accepted for gene ontology analysis. Here we show the top 30 gene ontology categories, classified by p-value. The most represented categories are related to skeletal muscle function, myofiber organization, morphology, and differentiation.

### 4.4 - Upstream Regulators

We used Upstream Regulator Analysis, a tool available in Ingenuity Pathway Analysis (IPA), to predict possible upstream regulators responsible for modulating the genes differentially regulated in COL6RD muscle biopsy samples. We considered only the upstream regulators with a z-score > 2 or <-2, and a p-value < 0.05. Based on the differentially expressed genes present in our dataset, TGFβ1 was most commonly identified regulator predicted as activated (p-value = 6.48E-18; z-score= 5.627), applying to the regulation of 47 downstream effectors genes, all of them being upregulated in COL6RD muscle biopsies. In the same list we also found downstream targets of TGFβ1 that in turn become regulators of other distinct subset of genes, such as CTGF (p-value = 1.77E-05; z-score= 2.402), indicating a multilevel complexity in the gene regulation specific to our cohort of COL6RD.

Among the upstream regulators predicted to be inhibited are estrogen receptor (ESR, p-value = 1.38E-15; z-score= -4.062), the transcription regulator SPDEF (p-value = 6.24E-14; z-score= -3.935), and the growth factor WISP2 (p-value = 2.91E-08; z-score= -2.714) (Table 7).

| Upstream Regulator | p-value | Activation z-score | Activation State | Upstream Regulator | p-value | Activation z-score | Activation State |
|---|---|---|---|---|---|---|---|
| TGFB1 | 6.48E-18 | 5.627 | Activated | HDAC6 | 7.63E-07 | 2.157 | Activated |
| ERK | 6.21E-06 | 3.553 | Activated | PGR | 1.30E-08 | 2.116 | Activated |
| ERBB2 | 1.78E-07 | 3.313 | Activated | EGFR | 1.71E-04 | 2.109 | Activated |
| SYVN1 | 2.29E-07 | 3.298 | Activated | IL5 | 9.10E-02 | 2.000 | Activated |
| P38 MAPK | 3.25E-07 | 3.146 | Activated | RLIM | 1.75E-04 | 2.000 | Activated |
| PDGF BB | 1.41E-12 | 3.112 | Activated | SMARCD3 | 1.59E-03 | 2.000 | Activated |
| Jnk | 5.30E-04 | 2.765 | Activated | MAP3K14 | 1.87E-02 | 2.000 | Activated |
| GLI1 | 3.24E-08 | 2.746 | Activated | TAZ | 5.14E-05 | 2.000 | Activated |
| Pkc(s) | 2.87E-04 | 2.574 | Activated | NEDD9 | 6.69E-03 | 2.000 | Activated |
| HIF1A | 4.18E-10 | 2.552 | Activated | RBPJ | 2.00E-03 | -2.000 | Inhibited |
| CCL5 | 3.11E-03 | 2.449 | Activated | MAP3K7 | 2.88E-02 | -2.000 | Inhibited |
| RELA | 1.11E-02 | 2.425 | Activated | miR-29b-3p | 2.38E-03 | -2.219 | Inhibited |
| PRKCD | 1.35E-03 | 2.414 | Activated | Hdac | 3.57E-05 | -2.273 | Inhibited |
| CTGF | 1.77E-05 | 2.402 | Activated | COL18A1 | 1.08E-04 | -2.538 | Inhibited |
| TWIST1 | 3.13E-04 | 2.236 | Activated | WISP2 | 2.91E-08 | -2.714 | Inhibited |
| EIF2AK2 | 4.87E-02 | 2.236 | Activated | FBN1 | 2.75E-09 | -2.797 | Inhibited |
| MYB | 1.20E-04 | 2.215 | Activated | FOXA1 | 1.08E-04 | -2.802 | Inhibited |
| Akt | 3.52E-03 | 2.213 | Activated | IgG | 2.25E-06 | -3.500 | Inhibited |
| TWIST2 | 1.26E-04 | 2.207 | Activated | MGEA5 | 1.30E-13 | -3.550 | Inhibited |
| CSF2 | 1.62E-01 | 2.200 | Activated | SPDEF | 6.24E-14 | -3.935 | Inhibited |
| LDL | 8.56E-03 | 2.186 | Activated | Estrogen Receptor | 1.38E-15 | -4.062 | Inhibited |

**Table 7. Upstream regulator analysis.**

Upstream regulator analysis performed on the list of differentially regulated genes obtained from the combination of the uniquely occurring genes from microarray and RNA-Seq. We show the top 30 activated upstream regulators, classified by p-value and z-score, and the top 12 inhibited upstream regulators.

### 4.5 - Validation of differentially expressed genes by qRT-PCR

We performed technical validation using quantitative reverse transcription-PCR (qRT-PCR) in order to validate expression level of selected entries among the most upregulated genes and to validate the involvement of specific molecular pathways of interest. We selected a series of 16 target genes (*BECN1, CILP, COL1A2, COL3A1, CTGF, FN1, LUM, MGP, MYH1, P21/CDKN1A, TGFβ1, THBS4, BGN, TP53, ATROGIN1/MAFBX1, BNIP3*) and one reference gene that showed low variance and was not differentially expressed between COL6RD and controls (TBP). We found high agreement between the relative gene expression levels measured with microarray and RNA-Seq, and qRT-PCR for all of the genes tested. However, the magnitude of the expression level between COL6RD and controls samples measured with qRT-PCR was always more pronounced. We confirmed the regulation of TGFβ1 pathway at the transcriptional level, with the increased expression of *COL1A2, COL3A1,CTGF, FN1*, and *BGN*. mRNA level of TGFβ1 resulted significantly increased when measured with qRT-PCR only. The upregulation of cartilage and bone formation related genes (*CILP, MGP, LUM*) was also confirmed. We explored further the autophagy pathway by measuring *BECN1* and *BNIP3*, key initiators of the autophagy process. The expression levels of *BECN1* and *BNIP3* were not different between controls and COL6RD muscle biopsies. We also confirmed that *ATROGIN1*, the E3 ubiquitin ligase involved in the FOXO3-driven atrophy program in skeletal muscle, was transcriptionally repressed (Figure 8).

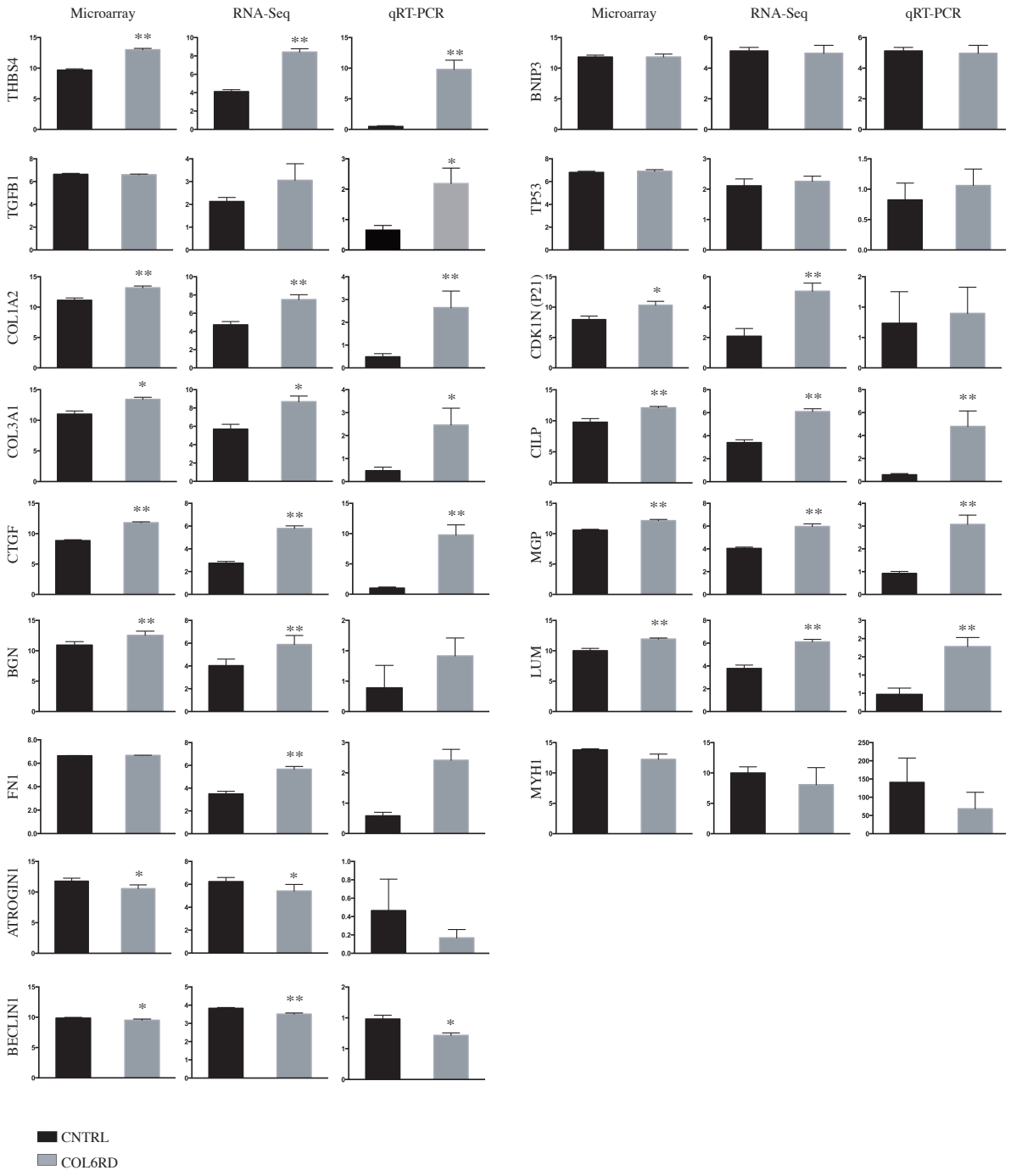



**Figure 8. qRT-PCR technical validation.**

Technical validation performed by qPCR on 16 genes performed on a random selection of n=5 controls (30%) and n=6 COL6RD (27%) muscle biopsies, accompanied by the post-noise filtering expression levels for the same samples detected by RNA-Seq and microarray. The relative expression levels are plotted on the y-axis ± S.E.M. The statistical significance was established using the non-parametric Mann-Whitney test. * indicates $p < 0.05$, ** indicates $p < 0.001$

### 4.6 - Validation of selected protein targets

We further performed protein validation of some of the products included in the first 20 most upregulated genes and somehow related to collagen VI, and Beclin1, Lc3b because previously implicated in the pathogenetic mechanisms of the Col6a1-/- mouse model of collagen VI-deficiency (Grumati et al., 2010). The western blotting analysis was performed on 4 controls and 9 COL6RD muscle biopsies. Thrombospondin-4, decorin and biglycan protein levels were variably upregulated in muscle biopsies. Beclin1 and Lc3b protein levels were variable throughout the sample pool, but generally not significantly different between controls and COL6RD.

TGFβ1 was predicted by the upstream regulator analysis performed with Ingenuity Pathway Analysis as activated in COL6RD muscle biopsies. To explore further this pathway we performed protein validation of the phospshorylated form of Smad2 and total Smad2. Our results showed that there were no increase in the level of phosphorylated-Smad2 in COL6RD muscle biopsies when compared to normal controls. The protein level of Smad2 were unchanged between normal controls and COL6RD muscle biopsies. (Figure 9, upper panel).

We further assessed the localization of THBS4 by immunofluorescence. We probed controls (n=3) and COL6RD muscle biopsies (n=4) with antibodies recognizing THBS4 and calnexin, an endoplasmic reticulum marker, and perlecan as skeletal muscle basal lamina marker. As the images show, THBS4 protein levels are upregulated in the patients muscle biopsies compared to controls. Also, while in controls THBS4 seems to be localized mostly around the endomysial capillaries, also identified by perlecan, and somewhat around the nuclei almost coinciding with calnexin staining, in the COL6RD muscle biopsies THBS4 localizes mostly along the periphery of the skeletal muscle fibers. Despite some level of colocalization with calnexin, the majority of THBS4 does not perfectly overlay with the ER stain (Figure 9 lower panel).

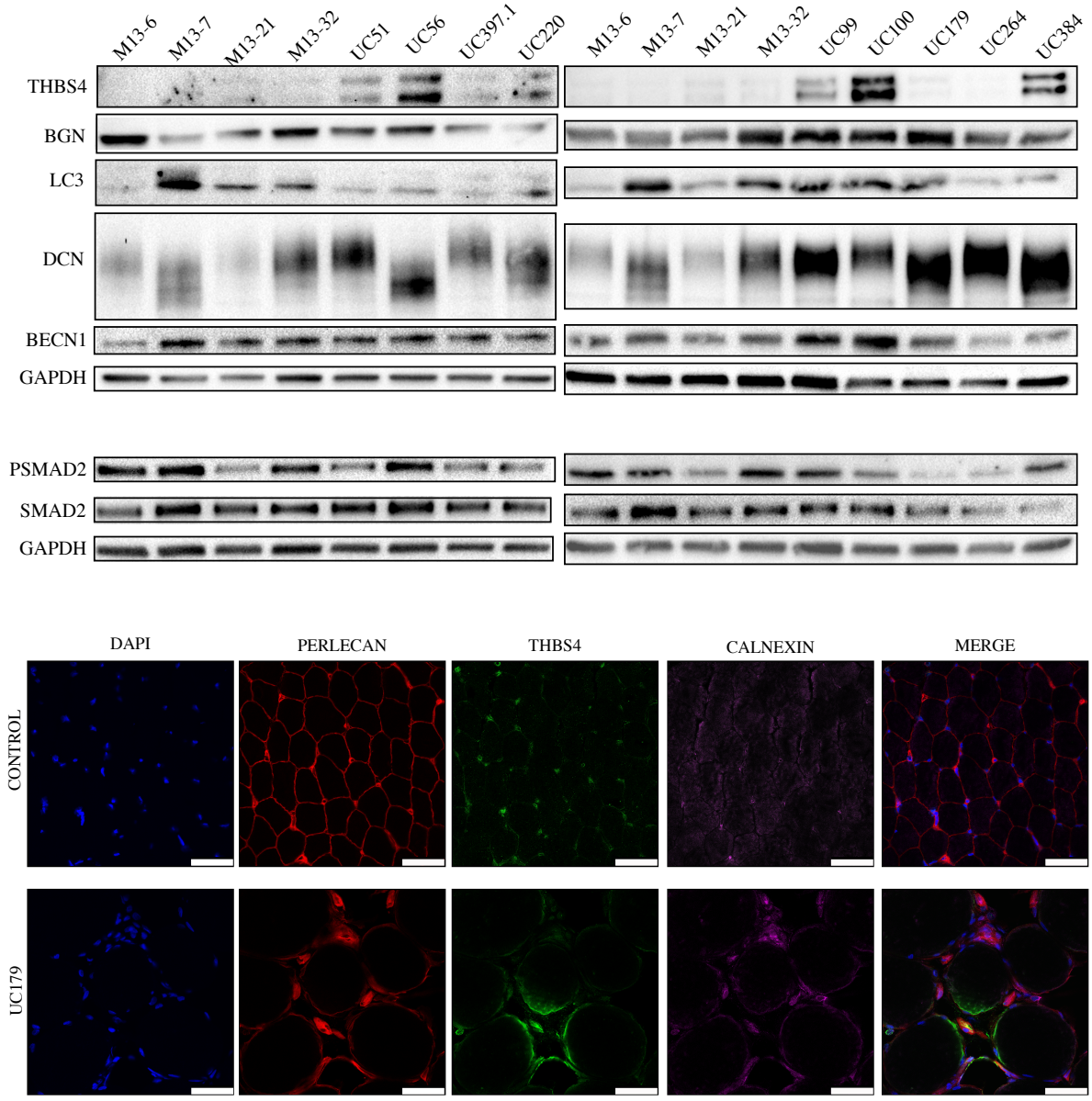



**Figure 9. Protein validation.**

Protein validation performed by western blot (upper) and immunofluorescence (lower) on controls (n=3) and COL6RD (n=3) muscle biopsies. Magnification 63X. The scale bar corresponds to 50 μm.

## 5 - DISCUSSION

The aim of the present study was to investigate pathophysiological pathways involved in the determination of the different histological states of the collagen VI-related muscular dystrophies (COL6RD). We performed microarray-based and RNA-Seq - based gene expression profiling in the same set of COL6RD muscle biopsies and normal controls to characterize the progression of COL6RD as reflected by relative gene expression and transcript patterns and we compared the two different approaches to expression profiling.

We also proposed a systematic review and histological classification of COL6RD muscle biopsies aided by an automated image analysis algorithm, along with two different approaches to gene expression profiling, microarray and RNA-Seq. We showed that different histological stages of COL6RD are characterized by the same pattern of gene expression with consequent no correlation between the histopathological changes in COL6RD muscle biopsies and the relative gene expression patterns. Rather, even in the early biopsies without histological evidence for fibrosis the extracellular matrix signature was already fully established. This observation helps to establish COL6RD as a primary disorder of the muscle extracellular matrix. Our results also suggested that TGFβ1 act as a major driver of the pathological changes at the skeletal muscle level. Moreover, muscle biopsies from COL6RD patients are characterized by an increase in alternative spliced isoforms of extracellular matrix-related genes that might contribute to the pathogenesis of the disease.

In skeletal muscle, collagen type VI is produced by the interstitial fibroblasts and it is thought to anchor the basement membrane in the skeletal muscle by interacting with collagen IV, a major component of the basal lamina, and other ECM component like biglycan and decorin (Bidanset et al., 1992; Kuo et al., 1997; Wiberg et al., 2001; Zou et al., 2008). Unbiased gene ontology analysis of the differentially regulated genes identified genes terms related to the extracellular matrix, all of them being upregulated in COL6RD muscle biopsies. Among the first 20 most upregulated genes in COL6RD muscle biopsies we identified fibril-forming collagens, such as COL1A2 and COL3A1, and the extracellular matrix proteoglycans biglycan, decorin, and lumican. This findings can be ascribed to two different processes. The first is the development of fibrosis, which is a prominent characteristics of muscle histology in COL6RD. In this case, the upregulation of extracellular matrix components would indicate increased synthesis of non functional connective tissue. These findings are common to other neuromuscular disorders, such as Duchenne muscular dystrophy, limb girdle muscular dystrophy 2A, Pompe disease (Bakay et al., 2002; Haslett et al.,

2002; Palermo et al., 2012; Saenz et al., 2008; Zanotti et al., 2005). The second reason for the upregulation of extracellular matrix component, particularly of biglycan and decorin, might be a tentative to reorganize and preserve the extracellular matrix-skeletal muscle interface due to the absence or mislocalization of collagen VI. Previous studies have shown that decorin and biglycan have the ability to bind collagen I and collagen VI, and biglycan can also bind α- and γ-sarcoglycan of the dystrophin-sarcoglycan complex in skeletal muscle (Rafii et al., 2006; Wiberg et al., 2001). When secreted in the extracellular matrix, they cooperate in the process of fibrillogenesis, especially in cartilage and tendons. Absence of decorin and biglycan appear to alter strength and stiffness of tendons, reinforcing their role in organizing the connective tissue (G. Zhang et al., 2006). Given the localization and the role that these small proteoglycans play in the extracellular matrix, it is possible that decorin and biglycan are being upregulated in response to the absence of collagen VI to rescue the function of the extracellular matrix, to preserve the anchorage of the muscle to the matrix and to preserve its biomechanic properties by acting on collagen I and other extracellular matrix proteins to reorganize the collagen fibrils, and to the basement membrane of the skeletal muscle to maintain the integrity of the system as a whole.

Interestingly, the upregulation of the extracellular matrix components was also present in the less severe group of COL6RD muscle biopsies, where there were no histological evidence of fibrosis yet. Inter individual variability due to different genetic backgrounds of the participants to this study might play a role in modulating the individual response to pathologic stimuli, together with environmental factors. In this specific case, genetic modifiers present in a subgroup of the patients population might have contributed to the rate at which the fibrosis progresses, despite the early activation of pro-fibrotic signals, delaying the histological appearance of fibrosis. At the molecular level, protein synthesis leading to connective tissue deposition might have overcome the protein degradation of the same extra connective tissue. Metalloproteinases (MMPs), $Zn^{2+}$ - dependent extracellular endopeptidases, are the major agents in charge of the protein degradation of extracellular components and play an important role in skeletal muscle physiology and pathology (Sbardella et al., 2015). Some MMPs are closely dependent on specific extracellular matrix partners to be fully activated and functional. In the specific case of COL6RD, it can be speculated that at the earlier stages of the disease, when the muscle is not severely compromised, the MMPs still function within their optimal level range, and they can still compensate for the pathological deposition of connective tissue by degrading it, despite the early raise of fibrosis-related genes. As the disease progresses, the muscle and the matrix become less and less functional, the MMPs might lose contact with functional partners or get locked in the matrix so they fail to compensate for the

deposition of ECM, and the protein synthesis of collagens overcome their degradation, leading to a progressive and degenerative fibrotic process. All of these hypothesis would required further targeted investigation to be proven or otherwise disproven.

The gene expression profiling performed in the present study identified an extensive network of genes that are known targets of the TGFβ/Smad2/3 axis, corroborated also by the upstream regulators analysis suggesting that TGFβ1 signaling is activated and upregulated in COL6RD muscle biopsies. Among the most upregulated genes regulated by TGFβ there are fibril-forming collagens (i.e. COL1A1, COL1A2, COL3A1), other extracellular matrix proteins (i.e. BGN, LTBP2, FN1), and integrins (ITGAV, ITGB2). However, although we assessed the gene and protein upregulation of downstream targets of TGFβ1, we did not observed any transcriptional regulation of the SMADs and, concordantly, we did not observed an increase in the phosphorylation of Smad2, effector of the TGFβ1 canonical pathway. This can be partially explained if we consider that in this particular case TGFβ1 might not act through the canonical Smad-dependent pathway, rather via the non canonical pathways that involve other key proteins, such as Erk, Jnk, and Rho-GTPase. The upstream regulator analysis performed with Ingenuity Pathway Analysis would support the activation of non canonical pathways for TGFβ1 (Y. E. Zhang, 2009). However, further studies are needed to gain insights about this process.

The most upregulated gene in our dataset was Thrombospondin-4 (THBS4). Expression of THBS4 is restricted to skeletal and cardiac muscle, and to the nervous system. In skeletal muscle, THBS4 is mainly expressed by interstitial cells and it is particularly abundant in the endomysium and perimysium of oxidative muscles (i.e. soleus), as well as in tendons, where it regulates the structure of the collagen fibrils (Arber & Caroni, 1995; Frolova et al., 2014). Although the function of THBS4 in skeletal muscle remains to be elucidated, it was recently reported that THBS4 might have a protective role in myocardial remodeling by regulating matrix deposition by fibroblasts and fibrosis independent of inflammation (Frolova et al., 2012). In the absence of functional data on THBS4 role in skeletal muscle, we can speculate that the increase in expression of THBS4 in COL6RD may reflect either an attempt to prevent or reduce the deposition of connective tissue therefore counteracting the fibrogenic process, or a yet not known contribution to the pathogenesis of COL6RD by aberrantly regulating the structure of the collagen fibrils. Recently, intracellular THBS4 has been shown to interact with the protein ATF6, sensor of misfolded protein in the ER and effector of the response to ER stress, in cardiomyocites (Lynch et al., 2012). In this study, THBS4 and other members of the Thrombospondins family would produce a protective ER stress response through ATF6 activation. Although we cannot exclude such a role in skeletal muscle, we did not

observe the upregulation of typical ER stress markers, and we did not observe a strict colocalization of THBS4 and the ER marker calnexin. Therefore, the role of THBS4 in ER stress response in skeletal muscle remains to be confirmed.

One finding from our gene expression profiling which appears to be specific to COL6RD, as it has not yet been reported in other muscular dystrophies, is the overexpression of cartilage specific genes. Cartilage intermediate layer protein (CILP) is a gene specifically expressed in articular cartilage and it is absent from non cartilaginous tissues. This protein is synthesized and deposited in the extracellular matrix by chondrocytes. Although the physiological function of the CILP protein remains largely unclear, it has been implicated in calcium crystal deposition in articular cartilage in patients with osteoarthritis and rheumatoid arthritis in response to age and increased concentration of TGFβ1 (Hirose, Masuda, & Ryan, 2000; Lorenzo, Bayliss, & Heinegard, 1998; Mori et al., 2006). To date, there is no literature reporting expression of CILP by myofibers, muscle interstitial cells (i.e. fibroblasts), or muscle derived mesenchymal stem cells. MGP is another gene upregulated in the muscle biopsies of COL6RD patients. It codes for a secreted vitamin K-dependent protein synthesized by a broad range of tissues and cells, including fibroblasts, chondrocytes, osteoblasts, and vascular endothelial cells. Despite the broad tissue distribution of MGP, only cartilage, bone, and calcified cartilage have shown to accumulate significant levels of MGP (Cancela, Hsieh, Francke, & Price, 1990). One reported function of MGP is the inhibition of calcification of soft connective tissues. MGP-deficient mice shows extensive calcification of the aorta and articular cartilage (Gheduzzi et al., 2007; Luo et al., 1997). Our study demonstrated an approximately 3-fold increase in MGP and an approximately 5-fold increased in CILP in COL6RD muscle biopsies compared to controls. Interestingly, MGP and CILP are reported to have opposite effects when secreted in the extracellular matrix.

It is well known that muscle derived mesenchymal stem cells and fibroblasts can transdifferentiate in other cell types, like chondrocytes, osteoblasts, adipocyte, and myofibroblasts if stimulated by specific growth factors (Agley, Rowlerson, Velloso, Lazarus, & Harridge, 2013; Alexeev et al., 2014). TGFβ1 is the major candidates for driving the transdifferentiation process. It was previously reported that mesenchymal stem cells can differentiate in chondrocytes if subjected to TGFβ1 stimulation (Johnstone, Hering, Caplan, Goldberg, & Yoo, 1998; Palmer et al., 2005). Moreover, TGFβ1 can promote transdifferentiation of myoblasts and fibroblasts in vitro in myofibroblasts through the Smad2/3-dependent upregulation of SK1 and S1P3 (Bruno et al., 2015; Cencetti, Bernacchioni, Nincheri, Donati, & Bruni, 2010; Wu, Zhang, Salmon, Lin, & Zehner, 2007). The level at which this upregulation occurs is still unknown. It might be caused by

deregulation of intracellular signals that fail to inhibit Smad2/3/4 action. In alternative, it might be caused by an increased biosynthesis of TGFβ1 protein, or a persistence of TGFβ1 caused by alterations in the ECM structure that entrap TGFβ1 preventing its degradation and making it available for longer periods to its receptor. Therefore, a possible explanation for the presence of cartilage and bone specific genes (i.e. CILP and MGP) in the muscle biopsies from COL6RD patients is that either fibroblasts from the muscle interstitium, or muscle derived mesenchymal stem cells transdifferentiated into other cell types due to an deregulation or mislocalization of key growth factors and critical signaling pathways (i.e. TGFβ1 canonical or non canonical pathway). In turn, this would cause the activation of signals intended to counteract the differentiation stimuli or the effects thereof, in this case carried out by MGP.

The modulation of gene expression is achieved by a diverse multitude of transcriptional and post-transcriptional mechanisms. Non-coding RNA-mediated gene regulation has catalyzed the attention being implicated in controlling physiological and diseased states in skeletal muscle , and being a potential therapeutic strategy for many neuromuscular disorders (Eisenberg, Alexander, & Kunkel, 2009; Muntoni & Wood, 2011). About a 40% of the differentially regulated transcripts identified by RNA-Seq in COL6RD were non protein coding, which leaves room for speculations implicating non coding RNA (i.e. lncdRNA, miRNA) in the modulation of gene expression in COL6RD. However, further investigation is needed to gain specific and detailed insight into the roles and functions of regulatory RNAs in COL6RD.

Microarray and RNA-Seq have been use before to compare transcriptomic datasets deriving from the same RNA preparation in many fields, including cancer and immunology (Xu et al., 2013; Zhao, Fung-Leung, Bittner, Ngo, & Liu, 2014). The general consensus is that there is strong concordance between microarray and RNA-Seq gene expression measurement across the platforms. This was confirmed also in the present study, where we show a correlation in both direction and magnitude of the fold change greater than 0.8. However, in our study, microarray detected more significantly differentially expressed genes when compared to RNA-Seq, which is generally the opposite of what happened in previous studies (Marioni, Mason, Mane, Stephens, & Gilad, 2008; Mortazavi, Williams, McCue, Schaeffer, & Wold, 2008). This might be due to the more stringent statistical analysis applied. For example, considering the normalized data prior to statistical testing, RNA-Seq and microarray detected the same genes. However, many genes detected by RNA-Seq failed to pass the false discovery rate threshold. The same phenomenon was observed with regards of the microarray for the FN1 gene, which was detected by RNA-Seq and validated by qPCR, but it failed to pass the statistical significance test when considering microarray.

Interestingly, we observed that roughly 30% of the total reads obtained by RNA-Seq for each sample were spread over only 24 genes, out of more than 57,000 known genes present in the human genome, with an average fold change magnitude lower than 1.5, therefore below the cutoff and not selected for further consideration. These 24 genes were represented mostly by different types of myosin, thus strictly skeletal muscle-specific genes, and GAPDH, abundantly present in any type of tissue. This introduces a major bias when performing RNA-Seq versus microarray. In fact, while microarray provide limited dynamic range, the coverage is evenly distributed among all the genes represented on the array. On the other hand, RNA-Seq provides virtually unlimited dynamic range, but coverage can result uneven in favor of overrepresented genes that are not necessarily differentially expressed among samples, but that would take over the available sequencing depth, therefore providing a major bias against poorly expressed genes. This is the reason why, in our study, RNA-Seq detected less genes than microarray in general (8234 of microarray vs 2889 of RNA-Seq), and of these fewer were statistically significant between COL6RD and controls.

Despite some bias dictated by the method, the technical validation analysis with qRT-PCR done on genes detected by both platforms suggested that both microarray and RNA-Seq were highly reliable and reproducible to detect gene expression differences.

With the present study, we intend to offer a transcriptomic approach to the study of COL6RD, in an effort to understand the molecular pathways responsible for the pathogenesis and the progression of the disease. Our results also suggest that TGFβ1 dependent pathways act as a major driver of the pathological changes at the skeletal muscle level in COL6RD, offering a new therapeutic approach to treat this specific subtype of neuromuscular disorder.

## 6 - BIBLIOGRAPHY


Agley, C. C., Rowlerson, A. M., Velloso, C. P., Lazarus, N. R., & Harridge, S. D. (2013). Human skeletal muscle fibroblasts, but not myogenic cells, readily undergo adipogenic differentiation. *J Cell Sci, 126*(Pt 24), 5610-5625. doi: 10.1242/jcs.132563

Alexeev, V., Arita, M., Donahue, A., Bonaldo, P., Chu, M. L., & Igoucheva, O. (2014). Human adipose-derived stem cell transplantation as a potential therapy for collagen VI-related congenital muscular dystrophy. *Stem Cell Res Ther, 5*(1), 21. doi: 10.1186/scrt411

Arber, S., & Caroni, P. (1995). Thrombospondin-4, an extracellular matrix protein expressed in the developing and adult nervous system promotes neurite outgrowth. *J Cell Biol, 131*(4), 1083-1094.

Arts, W. F., Bethlem, J., & Volkers, W. S. (1978). Further investigations on benign myopathy with autosomal dominant inheritance. *J Neurol, 217*(3), 201-206.

Bakay, M., Zhao, P., Chen, J., & Hoffman, E. P. (2002). A web-accessible complete transcriptome of normal human and DMD muscle. *Neuromuscul Disord, 12 Suppl 1*, S125-141.

Baker, N. L., Morgelin, M., Pace, R. A., Peat, R. A., Adams, N. E., Gardner, R. J., . . . Lamande, S. R. (2007). Molecular consequences of dominant Bethlem myopathy collagen VI mutations. *Ann Neurol, 62*(4), 390-405. doi: 10.1002/ana.21213

Baker, N. L., Morgelin, M., Peat, R., Goemans, N., North, K. N., Bateman, J. F., & Lamande, S. R. (2005). Dominant collagen VI mutations are a common cause of Ullrich congenital muscular dystrophy. *Hum Mol Genet, 14*(2), 279-293. doi: 10.1093/hmg/ddi025

Bethlem, J., & Wijngaarden, G. K. (1976). Benign myopathy, with autosomal dominant inheritance. A report on three pedigrees. *Brain, 99*(1), 91-100.

Bidanset, D. J., Guidry, C., Rosenberg, L. C., Choi, H. U., Timpl, R., & Hook, M. (1992). Binding of the proteoglycan decorin to collagen type VI. *J Biol Chem, 267*(8), 5250-5256.

Bruno, G., Cencetti, F., Pertici, I., Japtok, L., Bernacchioni, C., Donati, C., & Bruni, P. (2015). CTGF/CCN2 exerts profibrotic action in myoblasts via the up-regulation of sphingosine kinase-1/S1P3 signaling axis: Implications in the action mechanism of TGFbeta. *Biochim Biophys Acta, 1851*(2), 194-202. doi: 10.1016/j.bbalip.2014.11.011

Bruns, R. R. (1984). Beaded filaments and long-spacing fibrils: relation to type VI collagen. *J Ultrastruct Res, 89*(2), 136-145.

Butterfield, R. J., Foley, A. R., Dastgir, J., Asman, S., Dunn, D. M., Zou, Y., . . . Bonnemann, C. G. (2013). Position of glycine substitutions in the triple helix of COL6A1, COL6A2, and

COL6A3 is correlated with severity and mode of inheritance in collagen VI myopathies. *Hum Mutat, 34*(11), 1558-1567. doi: 10.1002/humu.22429

Cancela, L., Hsieh, C. L., Francke, U., & Price, P. A. (1990). Molecular structure, chromosome assignment, and promoter organization of the human matrix Gla protein gene. *J Biol Chem, 265*(25), 15040-15048.

Cencetti, F., Bernacchioni, C., Nincheri, P., Donati, C., & Bruni, P. (2010). Transforming growth factor-beta1 induces transdifferentiation of myoblasts into myofibroblasts via up-regulation of sphingosine kinase-1/S1P3 axis. *Mol Biol Cell, 21*(6), 1111-1124. doi: 10.1091/mbc.E09-09-0812

Chu, M. L., Mann, K., Deutzmann, R., Pribula-Conway, D., Hsu-Chen, C. C., Bernard, M. P., & Timpl, R. (1987). Characterization of three constituent chains of collagen type VI by peptide sequences and cDNA clones. *Eur J Biochem, 168*(2), 309-317.

Chu, M. L., Pan, T. C., Conway, D., Kuo, H. J., Glanville, R. W., Timpl, R., . . . Deutzmann, R. (1989). Sequence analysis of alpha 1(VI) and alpha 2(VI) chains of human type VI collagen reveals internal triplication of globular domains similar to the A domains of von Willebrand factor and two alpha 2(VI) chain variants that differ in the carboxy terminus. *EMBO J, 8*(7), 1939-1946.

Demir, E., Sabatelli, P., Allamand, V., Ferreiro, A., Moghadaszadeh, B., Makrelouf, M., . . . Guicheney, P. (2002). Mutations in COL6A3 cause severe and mild phenotypes of Ullrich congenital muscular dystrophy. *Am J Hum Genet, 70*(6), 1446-1458. doi: 10.1086/340608

Eisenberg, I., Alexander, M. S., & Kunkel, L. M. (2009). miRNAS in normal and diseased skeletal muscle. *J Cell Mol Med, 13*(1), 2-11. doi: 10.1111/j.1582-4934.2008.00524.x

Engel, J., Furthmayr, H., Odermatt, E., von der Mark, H., Aumailley, M., Fleischmajer, R., & Timpl, R. (1985). Structure and macromolecular organization of type VI collagen. *Ann N Y Acad Sci, 460*, 25-37.

Fitzgerald, J., Rich, C., Zhou, F. H., & Hansen, U. (2008). Three novel collagen VI chains, alpha4(VI), alpha5(VI), and alpha6(VI). *J Biol Chem, 283*(29), 20170-20180. doi: 10.1074/jbc.M710139200

Foley, A. R., Hu, Y., Zou, Y., Columbus, A., Shoffner, J., Dunn, D. M., . . . Bonnemann, C. G. (2009). Autosomal recessive inheritance of classic Bethlem myopathy. *Neuromuscul Disord, 19*(12), 813-817. doi: 10.1016/j.nmd.2009.09.010

Frolova, E. G., Drazba, J., Krukovets, I., Kostenko, V., Blech, L., Harry, C., . . . Stenina-Adognravi, O. (2014). Control of organization and function of muscle and tendon by thrombospondin-4. *Matrix Biol, 37*, 35-48. doi: 10.1016/j.matbio.2014.02.003

Frolova, E. G., Sopko, N., Blech, L., Popovic, Z. B., Li, J., Vasanji, A., . . . Stenina, O. I. (2012). Thrombospondin-4 regulates fibrosis and remodeling of the myocardium in response to pressure overload. *FASEB J, 26*(6), 2363-2373. doi: 10.1096/fj.11-190728

Furukawa, T., & Toyokura, Y. (1977). Congenital, hypotonic-sclerotic muscular dystrophy. *J Med Genet, 14*(6), 426-429.

Gara, S. K., Grumati, P., Urciuolo, A., Bonaldo, P., Kobbe, B., Koch, M., . . . Wagener, R. (2008). Three novel collagen VI chains with high homology to the alpha3 chain. *J Biol Chem, 283*(16), 10658-10670. doi: 10.1074/jbc.M709540200

Garber, M., Grabherr, M. G., Guttman, M., & Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods, 8*(6), 469-477. doi: 10.1038/nmeth.1613

Gheduzzi, D., Boraldi, F., Annovi, G., DeVincenzi, C. P., Schurgers, L. J., Vermeer, C., . . . Ronchetti, I. P. (2007). Matrix Gla protein is involved in elastic fiber calcification in the dermis of pseudoxanthoma elasticum patients. *Lab Invest, 87*(10), 998-1008. doi: 10.1038/labinvest.3700667

Giusti, B., Lucarini, L., Pietroni, V., Lucioli, S., Bandinelli, B., Sabatelli, P., . . . Pepe, G. (2005). Dominant and recessive COL6A1 mutations in Ullrich scleroatonic muscular dystrophy. *Ann Neurol, 58*(3), 400-410. doi: 10.1002/ana.20586

Grumati, P., Coletto, L., Sabatelli, P., Cescon, M., Angelin, A., Bertaggia, E., . . . Bonaldo, P. (2010). Autophagy is defective in collagen VI muscular dystrophies, and its reactivation rescues myofiber degeneration. *Nat Med, 16*(11), 1313-1320. doi: 10.1038/nm.2247

Gualandi, F., Urciuolo, A., Martoni, E., Sabatelli, P., Squarzoni, S., Bovolenta, M., . . . Merlini, L. (2009). Autosomal recessive Bethlem myopathy. *Neurology, 73*(22), 1883-1891. doi: 10.1212/WNL.0b013e3181c3fd2a

Haslett, J. N., Sanoudou, D., Kho, A. T., Bennett, R. R., Greenberg, S. A., Kohane, I. S., . . . Kunkel, L. M. (2002). Gene expression comparison of biopsies from Duchenne muscular dystrophy (DMD) and normal skeletal muscle. *Proc Natl Acad Sci U S A, 99*(23), 15000-15005. doi: 10.1073/pnas.192571199

Hirose, J., Masuda, I., & Ryan, L. M. (2000). Expression of cartilage intermediate layer protein/nucleotide pyrophosphohydrolase parallels the production of extracellular inorganic pyrophosphate in response to growth factors and with aging. *Arthritis Rheum, 43*(12), 2703-2711. doi: 10.1002/1529-0131(200012)43:12<2703::AID-ANR10>3.0.CO;2-Y

Irwin, W. A., Bergamin, N., Sabatelli, P., Reggiani, C., Megighian, A., Merlini, L., . . . Bonaldo, P. (2003). Mitochondrial dysfunction and apoptosis in myopathic mice with collagen VI deficiency. *Nat Genet, 35*(4), 367-371. doi: 10.1038/ng1270

Jobsis, G. J., Keizers, H., Vreijling, J. P., de Visser, M., Speer, M. C., Wolterman, R. A., . . . Bolhuis, P. A. (1996). Type VI collagen mutations in Bethlem myopathy, an autosomal dominant myopathy with contractures. *Nat Genet, 14*(1), 113-115. doi: 10.1038/ng0996-113

Johnstone, B., Hering, T. M., Caplan, A. I., Goldberg, V. M., & Yoo, J. U. (1998). In vitro chondrogenesis of bone marrow-derived mesenchymal progenitor cells. *Exp Cell Res, 238*(1), 265-272. doi: 10.1006/excr.1997.3858

Kuo, H. J., Keene, D. R., & Glanville, R. W. (1995). The macromolecular structure of type-VI collagen. Formation and stability of filaments. *Eur J Biochem, 232*(2), 364-372.

Kuo, H. J., Maslen, C. L., Keene, D. R., & Glanville, R. W. (1997). Type VI collagen anchors endothelial basement membranes by interacting with type IV collagen. *J Biol Chem, 272*(42), 26522-26529.

Lamande, S. R., Bateman, J. F., Hutchison, W., McKinlay Gardner, R. J., Bower, S. P., Byrne, E., & Dahl, H. H. (1998). Reduced collagen VI causes Bethlem myopathy: a heterozygous COL6A1 nonsense mutation results in mRNA decay and functional haploinsufficiency. *Hum Mol Genet, 7*(6), 981-989.

Lamande, S. R., Morgelin, M., Selan, C., Jobsis, G. J., Baas, F., & Bateman, J. F. (2002). Kinked collagen VI tetramers and reduced microfibril formation as a result of Bethlem myopathy and introduced triple helical glycine mutations. *J Biol Chem, 277*(3), 1949-1956. doi: 10.1074/jbc.M109932200

Lamande, S. R., Shields, K. A., Kornberg, A. J., Shield, L. K., & Bateman, J. F. (1999). Bethlem myopathy and engineered collagen VI triple helical deletions prevent intracellular multimer assembly and protein secretion. *J Biol Chem, 274*(31), 21817-21822.

Lamande, S. R., Sigalas, E., Pan, T. C., Chu, M. L., Dziadek, M., Timpl, R., & Bateman, J. F. (1998). The role of the alpha3(VI) chain in collagen VI assembly. Expression of an alpha3(VI) chain lacking N-terminal modules N10-N7 restores collagen VI assembly, secretion, and matrix deposition in an alpha3(VI)-deficient cell line. *J Biol Chem, 273*(13), 7423-7430.

Lampe, A. K., Zou, Y., Sudano, D., O'Brien, K. K., Hicks, D., Laval, S. H., . . . Bonnemann, C. G. (2008). Exon skipping mutations in collagen VI are common and are predictive for severity and inheritance. *Hum Mutat, 29*(6), 809-822. doi: 10.1002/humu.20704

Leyten, Q. H., Gabreels, F. J., Joosten, E. M., Renier, W. O., Ter Laak, H. J., Ter Haar, B. G., & Stadhouders, A. M. (1986). An autosomal dominant type of congenital muscular dystrophy. *Brain Dev, 8*(5), 533-537.

Lorenzo, P., Bayliss, M. T., & Heinegard, D. (1998). A novel cartilage protein (CILP) present in the mid-zone of human articular cartilage increases with age. *J Biol Chem, 273*(36), 23463-23468.

Lucarini, L., Giusti, B., Zhang, R. Z., Pan, T. C., Jimenez-Mallebrera, C., Mercuri, E., . . . Chu, M. L. (2005). A homozygous COL6A2 intron mutation causes in-frame triple-helical deletion and nonsense-mediated mRNA decay in a patient with Ullrich congenital muscular dystrophy. *Hum Genet, 117*(5), 460-466. doi: 10.1007/s00439-005-1318-8

Luo, G., Ducy, P., McKee, M. D., Pinero, G. J., Loyer, E., Behringer, R. R., & Karsenty, G. (1997). Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein. *Nature, 386*(6620), 78-81. doi: 10.1038/386078a0

Lynch, J. M., Maillet, M., Vanhoutte, D., Schloemer, A., Sargent, M. A., Blair, N. S., . . . Molkentin, J. D. (2012). A thrombospondin-dependent pathway for a protective ER stress response. *Cell, 149*(6), 1257-1268. doi: 10.1016/j.cell.2012.03.050

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., & Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res, 18*(9), 1509-1517. doi: 10.1101/gr.079558.108

Mercuri, E., Cini, C., Counsell, S., Allsop, J., Zolkipli, Z., Jungbluth, H., . . . Muntoni, F. (2002). Muscle MRI findings in a three-generation family affected by Bethlem myopathy. *Eur J Paediatr Neurol, 6*(6), 309-314.

Mercuri, E., Cini, C., Pichiecchio, A., Allsop, J., Counsell, S., Zolkipli, Z., . . . Muntoni, F. (2003). Muscle magnetic resonance imaging in patients with congenital muscular dystrophy and Ullrich phenotype. *Neuromuscul Disord, 13*(7-8), 554-558.

Mercuri, E., Lampe, A., Allsop, J., Knight, R., Pane, M., Kinali, M., . . . Muntoni, F. (2005). Muscle MRI in Ullrich congenital muscular dystrophy and Bethlem myopathy. *Neuromuscul Disord, 15*(4), 303-310. doi: 10.1016/j.nmd.2005.01.004

Merlini, L., Angelin, A., Tiepolo, T., Braghetta, P., Sabatelli, P., Zamparelli, A., . . . Bernardi, P. (2008). Cyclosporin A corrects mitochondrial dysfunction and muscle apoptosis in patients with collagen VI myopathies. *Proc Natl Acad Sci U S A, 105*(13), 5225-5229. doi: 10.1073/pnas.0800962105

Merlini, L., Morandi, L., Granata, C., & Ballestrazzi, A. (1994). Bethlem myopathy: early-onset benign autosomal dominant myopathy with contractures. Description of two new families. *Neuromuscul Disord, 4*(5-6), 503-511.

Mohire, M. D., Tandan, R., Fries, T. J., Little, B. W., Pendlebury, W. W., & Bradley, W. G. (1988). Early-onset benign autosomal dominant limb-girdle myopathy with contractures (Bethlem myopathy). *Neurology, 38*(4), 573-580.

Mori, M., Nakajima, M., Mikami, Y., Seki, S., Takigawa, M., Kubo, T., & Ikegawa, S. (2006). Transcriptional regulation of the cartilage intermediate layer protein (CILP) gene. *Biochem Biophys Res Commun, 341*(1), 121-127. doi: 10.1016/j.bbrc.2005.12.159

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods, 5*(7), 621-628. doi: 10.1038/nmeth.1226

Muntoni, F., & Wood, M. J. (2011). Targeting RNA to treat neuromuscular disease. *Nat Rev Drug Discov, 10*(8), 621-637. doi: 10.1038/nrd3459

Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet, 12*(2), 87-98. doi: 10.1038/nrg2934

Paco, S., Kalko, S. G., Jou, C., Rodriguez, M. A., Corbera, J., Muntoni, F., . . . Jimenez-Mallebrera, C. (2013). Gene expression profiling identifies molecular pathways associated with collagen VI deficiency and provides novel therapeutic targets. *PLoS One, 8*(10), e77430. doi: 10.1371/journal.pone.0077430

Palermo, A. T., Palmer, R. E., So, K. S., Oba-Shinjo, S. M., Zhang, M., Richards, B., . . . Pomponio, R. J. (2012). Transcriptional response to GAA deficiency (Pompe disease) in infantile-onset patients. *Mol Genet Metab, 106*(3), 287-300. doi: 10.1016/j.ymgme.2012.05.004

Palmer, G. D., Steinert, A., Pascher, A., Gouze, E., Gouze, J. N., Betz, O., . . . Ghivizzani, S. C. (2005). Gene-induced chondrogenesis of primary mesenchymal stem cells in vitro. *Mol Ther, 12*(2), 219-228. doi: 10.1016/j.ymthe.2005.03.024

Palstra, A. P., Beltran, S., Burgerhout, E., Brittijn, S. A., Magnoni, L. J., Henkel, C. V., . . . Planas, J. V. (2013). Deep RNA sequencing of the skeletal muscle transcriptome in swimming fish. *PLoS One, 8*(1), e53171. doi: 10.1371/journal.pone.0053171

Pan, T. C., Zhang, R. Z., Pericak-Vance, M. A., Tandan, R., Fries, T., Stajich, J. M., . . . Speer, M. C. (1998). Missense mutation in a von Willebrand factor type A domain of the alpha 3(VI) collagen gene (COL6A3) in a family with Bethlem myopathy. *Hum Mol Genet, 7*(5), 807-812.

Pan, T. C., Zhang, R. Z., Sudano, D. G., Marie, S. K., Bonnemann, C. G., & Chu, M. L. (2003). New molecular mechanism for Ullrich congenital muscular dystrophy: a heterozygous in-frame deletion in the COL6A1 gene causes a severe phenotype. *Am J Hum Genet, 73*(2), 355-369. doi: 10.1086/377107

Pepe, G., Bertini, E., Giusti, B., Brunelli, T., Comeglio, P., Saitta, B., . . . Abbate, R. (1999). A novel de novo mutation in the triple helix of the COL6A3 gene in a two-generation Italian family affected by Bethlem myopathy. A diagnostic approach in the mutations' screening of type VI collagen. *Neuromuscul Disord, 9*(4), 264-271.

Pepe, G., Giusti, B., Bertini, E., Brunelli, T., Saitta, B., Comeglio, P., . . . Chu, M. L. (1999). A heterozygous splice site mutation in COL6A1 leading to an in-frame deletion of the alpha1(VI) collagen chain in an italian family affected by bethlem myopathy. *Biochem Biophys Res Commun, 258*(3), 802-807. doi: 10.1006/bbrc.1999.0680

Pepe, G., Lucarini, L., Zhang, R. Z., Pan, T. C., Giusti, B., Quijano-Roy, S., . . . Chu, M. L. (2006). COL6A1 genomic deletions in Bethlem myopathy and Ullrich muscular dystrophy. *Ann Neurol, 59*(1), 190-195. doi: 10.1002/ana.20705

Porter, J. D., Khanna, S., Kaminski, H. J., Rao, J. S., Merriam, A. P., Richmonds, C. R., . . . Andrade, F. H. (2002). A chronic inflammatory response dominates the skeletal muscle molecular signature in dystrophin-deficient mdx mice. *Hum Mol Genet, 11*(3), 263-272.

Rafii, M. S., Hagiwara, H., Mercado, M. L., Seo, N. S., Xu, T., Dugan, T., . . . Fallon, J. R. (2006). Biglycan binds to alpha- and gamma-sarcoglycan and regulates their expression during development. *J Cell Physiol, 209*(2), 439-447. doi: 10.1002/jcp.20740

Ricci, E., Bertini, E., Boldrini, R., Sabatelli, M., Servidei, S., Mazziotta, M. R., . . . Tonali, P. (1988). Late onset scleroatonic familial myopathy (Ullrich disease): a study of two sibs. *Am J Med Genet, 31*(4), 933-942. doi: 10.1002/ajmg.1320310428

Saenz, A., Azpitarte, M., Armananzas, R., Leturcq, F., Alzualde, A., Inza, I., . . . Lopez de Munain, A. (2008). Gene expression profiling in limb-girdle muscular dystrophy 2A. *PLoS One, 3*(11), e3750. doi: 10.1371/journal.pone.0003750

Saitta, B., Timpl, R., & Chu, M. L. (1992). Human alpha 2(VI) collagen gene. Heterogeneity at the 5'-untranslated region generated by an alternate exon. *J Biol Chem, 267*(9), 6188-6196.

Saitta, B., Wang, Y. M., Renkart, L., Zhang, R. Z., Pan, T. C., Timpl, R., & Chu, M. L. (1991). The exon organization of the triple-helical coding regions of the human alpha 1(VI) and alpha 2(VI) collagen genes is highly similar. *Genomics, 11*(1), 145-153.

Sasaki, T., Hohenester, E., Zhang, R. Z., Gotta, S., Speer, M. C., Tandan, R., . . . Chu, M. L. (2000). A Bethlem myopathy Gly to Glu mutation in the von Willebrand factor A domain N2 of the collagen alpha3(VI) chain interferes with protein folding. *FASEB J, 14*(5), 761-768.

Sbardella, D., Sciandra, F., Gioia, M., Marini, S., Gori, A., Giardina, B., . . . Bozzi, M. (2015). alpha-dystroglycan is a potential target of matrix metalloproteinase MMP-2. *Matrix Biol, 41*, 2-7. doi: 10.1016/j.matbio.2014.11.007

Scacheri, P. C., Gillanders, E. M., Subramony, S. H., Vedanarayanan, V., Crowe, C. A., Thakore, N., . . . Hoffman, E. P. (2002). Novel mutations in collagen VI genes: expansion of the Bethlem myopathy phenotype. *Neurology, 58*(4), 593-602.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., . . . Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol, 7*, 3. doi: 10.1186/1471-2199-7-3

Stokes, D. G., Saitta, B., Timpl, R., & Chu, M. L. (1991). Human alpha 3(VI) collagen gene. Characterization of exons coding for the amino-terminal globular domain and alternative splicing in normal and tumor cells. *J Biol Chem, 266*(13), 8626-8633.

Trikka, D., Davis, T., Lapenta, V., Brahe, C., & Kessling, A. M. (1997). Human COL6A1: genomic characterization of the globular domains, structural and evolutionary comparison with COL6A2. *Mamm Genome, 8*(5), 342-345.

Vanegas, O. C., Zhang, R. Z., Sabatelli, P., Lattanzi, G., Bencivenga, P., Giusti, B., . . . Pepe, G. (2002). Novel COL6A1 splicing mutation in a family affected by mild Bethlem myopathy. *Muscle Nerve, 25*(4), 513-519.

Wiberg, C., Hedbom, E., Khairullina, A., Lamande, S. R., Oldberg, A., Timpl, R., . . . Heinegard, D. (2001). Biglycan and decorin bind close to the n-terminal region of the collagen VI triple helix. *J Biol Chem, 276*(22), 18947-18952. doi: 10.1074/jbc.M100625200

Wu, Y., Zhang, X., Salmon, M., Lin, X., & Zehner, Z. E. (2007). TGFbeta1 regulation of vimentin gene expression during differentiation of the C2C12 skeletal myogenic cell line requires Smads, AP-1 and Sp1 family members. *Biochim Biophys Acta, 1773*(3), 427-439. doi: 10.1016/j.bbamcr.2006.11.017

Xu, X., Zhang, Y., Williams, J., Antoniou, E., McCombie, W. R., Wu, S., . . . Li, E. (2013). Parallel comparison of Illumina RNA-Seq and Affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated HT-29 colon cancer cells and simulated datasets. *BMC Bioinformatics, 14 Suppl 9*, S1. doi: 10.1186/1471-2105-14-S9-S1

Zanotti, S., Negri, T., Cappelletti, C., Bernasconi, P., Canioni, E., Di Blasi, C., . . . Mora, M. (2005). Decorin and biglycan expression is differentially altered in several muscular dystrophies. *Brain, 128*(Pt 11), 2546-2555. doi: 10.1093/brain/awh635

Zhang, G., Ezura, Y., Chervoneva, I., Robinson, P. S., Beason, D. P., Carine, E. T., . . . Birk, D. E. (2006). Decorin regulates assembly of collagen fibrils and acquisition of biomechanical properties during tendon development. *J Cell Biochem, 98*(6), 1436-1449. doi: 10.1002/jcb.20776

Zhang, R. Z., Sabatelli, P., Pan, T. C., Squarzoni, S., Mattioli, E., Bertini, E., . . . Chu, M. L. (2002). Effects on collagen VI mRNA stability and microfibrillar assembly of three COL6A2 mutations in two families with Ullrich congenital muscular dystrophy. *J Biol Chem, 277*(46), 43557-43564. doi: 10.1074/jbc.M207696200

Zhang, Y. E. (2009). Non-Smad pathways in TGF-beta signaling. *Cell Res, 19*(1), 128-139. doi: 10.1038/cr.2008.328

Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One, 9*(1), e78644. doi: 10.1371/journal.pone.0078644

Zhou, X., Dimachkie, M. M., Xiong, M., Tan, F. K., & Arnett, F. C. (2004). cDNA microarrays reveal distinct gene expression clusters in idiopathic inflammatory myopathies. *Med Sci Monit, 10*(7), BR191-197.

Zou, Y., Zhang, R. Z., Sabatelli, P., Chu, M. L., & Bonnemann, C. G. (2008). Muscle interstitial fibroblasts are the main source of collagen VI synthesis in skeletal muscle: implications for congenital muscular dystrophy types Ullrich and Bethlem. *J Neuropathol Exp Neurol, 67*(2), 144-154. doi: 10.1097/nen.0b013e3181634ef7