UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DOTTORATO DI RICERCA IN
GENETICA MOLECOLARE, BIOTECNOLOGIE E MEDICINA SPERIMENTALE

settore scientifico disciplinare BIO/18 - Genetica

CICLO XXXIV

Identification and characterization of a novel vacuolar myopathy caused by a mutation in the *PLIN4* gene and possible strategies for therapy development.

Relatore:
Chiar.mo Prof. Giuseppe Borsani

Coordinatore del Dottorato:
Chiar.mo Prof. Eugenio Monti

Dottoranda:
dott.ssa Alessandra Ruggieri

Anno accademico 2020/2021

Table of Contents

Abstract

English abstract

In a previous report, a large Italian family with 19 affected individuals across four generations, was described as characterized by an adult-onset autosomal dominant vacuolar myopathy, with distinctive myopathology featuring rimmed ubiquitin-positive autophagic vacuolation [Di Blasi_2004]. No genetic cause was identified, but a 8.12 Mb common haplotype was determined on the short arm of chromosome 19 (19p13.3).

My first approach to the study of this family, focused on the refinement of the disease locus, thanks to a combined microsatellite and SNP based array analysis. Accordingly, I was able to narrow the common diseased haplotype to 5.12 Mb in the same region of chromosome 19.

We then applied a next generation sequencing (NGS) approach, starting with exome (performed on two affected patients) and genome analysis (on two affected patients and two healthy siblings). No rare likely pathogenic coding variants were present in the affected but not in the healthy siblings, therefore we decided to use RNA-Seq in order to better understand the possible functional implications of the many intronic and intergenic variants that were identified with the NGS techniques.

In the meantime, a young individual, daughter of an affected patient and presenting with a genetic recombination on the common haplotype, underwent neurological examination that revealed on the Medical Research Council (MRC) scale, weakness of the hand extensor (4/5), neck flexor (4/5) and hallux extensor (4/5) muscles. Moreover, her electromyography showed positive repetitive complex discharges at needle insertion and myopathic pattern, particularly in tibialis anterior and biceps brachialis muscles, a typical pattern present in the affected relatives.

She also agreed to needle muscle biopsy on which we performed immunohistochemical analysis, focusing our attention on p62/SQSTM1 and ubiquitinated proteins (FK2), two players of the pathway responsible for degradation of misfolded and aggregated proteins. We observed the co-localization of the two selected antibodies, in the subsarcolemmal region of the fibers as well as within the vacuoles.

Comparing the staining in this asymptomatic patient with the results obtained in the milder affected and severely affected individuals, we noticed that the positivity was remarkably following the severity of the clinical phenotype. In particular, the most affected patients showed nearly complete staining of all the fibers and vacuoles while the asymptomatic only has few affected fibers. The milder phenotypes had intermediate number of positive fibers and vacuoles.

We then reasoned that the protein that was so specifically marked by p62/SQSTM1 and FK2 antibody should have been our diseased protein.

Accordingly, thanks to a collaboration, we performed laser microdissection and mass spectrometry, on the vacuole content and we found that the only protein accumulating within the vacuoles and encoded by a gene present in the haplotype was perilipin-4 encoded by *PLIN4* gene.
The re-analysis of the NGS data, showed the presence of an unusual coverage's peak that was not present in the healthy siblings in a region of *PLIN4* gene, suggesting the presence of a possible repeat.
Indeed, long-read sequencing with the Oxford Nanopore Technology, identified a 9×99-mer expansion of the sequence encoding perilipin 4's amphipathic region encoded by exon 3 of *PLIN4* gene. This region is structurally related to the domains in α-synuclein and apolipoproteins that mediate both lipid binding and proneness to aggregation.
In order to identify the pathogenic mechanism leading to this disorder, we performed immunohistochemical analysis focusing on the specialized autophagic pathway responsible for the clearance of protein aggregates, defined aggrephagy.
We then confirmed that perilipin 4 accumulation in our patients' muscles is associated with activation of the ubiquitination-mediated aggrephagy pathway involved in autophagic clearance of aggregating proteins.
This new disease represents a protein aggregation disorder caused by genetic expansion of the aggregating protein segment, and apparent insufficiency of aggrephagy to control the accumulation.
The study of the molecular mechanisms at the cellular level, in the myoblasts cells derived from the patients' biopsies proved to be not feasible due to the absence of perilipin 4 expression in these cells.
Therefore, we decided to produce a cellular model expressing the wild-type or the mutant protein to be able to test the propensity of the MUT allele to form aggregates and to be able to study the molecular mechanisms thus involved.
Moreover, this cellular model could be used as a possible tool for screening different compounds that might act on the formation or clearance of these protein aggregates.

Abstract in italiano

Una grande famiglia con 19 individui affetti in quattro generazioni è stata descritta dal gruppo della dott.ssa Mora [Di Blasi et al, 2004] come caratterizzata da miopatia vacuolare dominante con insorgenza nell'età adulta e progressione lenta con vacuoli rimmed positivi all'ubiquitina.

La causa genetica non era stata all'epoca individuata, ma era stato identificato un aplotipo comune agli affetti in una regione di 8.12 Mb nel braccio corto del cromosoma 19 (19p13.3).

Durante il mio primo approccio allo studio di questa famiglia, mi sono focalizzata sulla miglior definizione dell'aplotipo grazie all'uso combinato di microsatelliti e array basati su SNP. Ho quindi potuto restringere a 5.12 Mb la regione condivisa dagli affetti della famiglia.

Abbiamo in seguito sequenziato l'esoma (di due pazienti) ed il genoma (di quattro pazienti, due affetti e due sani) utilizzando tecnologie di sequenziamento di nuova generazione (NGS). Non essendo state evidenziate nelle regioni codificanti varianti verosimilmente patogeniche presenti in tutti gli affetti ed assenti nei familiari sani, abbiamo deciso di ricorrere al sequenziamento dell'RNA per poter valutare l'eventuale impatto funzionale delle numerose varianti introniche ed intergeniche individuate con il sequenziamento del genoma.

Nel frattempo, una giovane paziente con una ricombinazione nella regione dell'aplotipo e figlia di una paziente con fenotipo conclamato, ha effettuato una visita neurologica che ha rivelato la presenza di una minima riduzione della forza degli estensori della mano (4/5), flessori del collo (4/5) ed estensori dell'alluce (4/5) secondo la scala del Medical Research Council (MRC). Inoltre, l'elettromiografia ha mostrato scariche positive ripetute complesse ed un pattern miopatico, particolarmente nel tibiale anteriore e bicipite, EMG tipico degli affetti nella famiglia. La paziente si è inoltre sottoposta a biopsia muscolare sulla quale abbiamo effettuato l'analisi immunoistochimica focalizzando la nostra attenzione su due proteine, p62/SQSTM1 e le proteine ubiqutiinate (FK2), che fanno parte del pathway responsabile della degradazione delle proteine *misfolded* ed aggregate. Questa analisi ha mostrato la colocalizzazione dei due anticorpi nella regione subsarcolemmale e all'interno dei vacuoli.

Effettuando lo stesso studio sui pazienti severi o mediamente affetti, abbiamo notato che l'intesità della marcatura era direttamente correlata alla severità del fenotipo. Nello specifico, i pazienti più severi avevano virtualmente tutte le fibre ed i vacuoli positivi a questi anticorpi, mentre la paziente asintomatica aveva solo poche fibre positive. I fenotipi intermedi mostravano un intermedio numero di fibre marcate.

Abbiamo quindi ipotizzato che la proteina che veniva così specificatamente ubiquitinata e riconosciuta da p62/SQSTM1, potesse essere la nostra proteina mutata.
Grazie ad una collaborazione, abbiamo quindi effettuato una microdissezione laser dei vacuoli, con successiva analisi del contenuto mediante spettrometria di massa, evidenziando così che la proteina accumulata nei vacuoli e codificata dal gene *PLIN4* presente sull'aplotipo comune, era perilipina 4.
Rianalizzando I dati di NGS abbiamo individuato un picco inusuale della copertura nel gene *PLIN4*, picco che non era presente nei fratelli sani. Questo picco era suggestivo di una possibile anomala ripetizione di una sequenza del gene stesso, ripetizione non direttamente individuabile dall'analisi delle sequenze ottenute mediante NGS.

Abbiamo pertanto effettuato un sequenziamento *long-read* usando la tecnologia Oxford Nanopore Technology, identificando così una espansione di 9x99 nucleotidi nella regione dell'esone 3, codificante per il dominio anfipatico di perilipina-4. Questa regione è strutturalmente correlata alle eliche anfipatiche presenti in α-sinucleina e nelle apolipoproteine che mediano il binding con i lipidi e sono soggette ad aggregazione.
Per poter identificare il meccanismo patogenetico sotteso a questa patologia, abbiamo effettuato un'analisi immunoistochimica focalizzandoci sul pathway dell'aggrefagia, cioè la rimozione di aggregati proteici tramite degradazione lisosomiale.
Abbiamo quindi confermato che l'accumulo di perilipina-4 nel muscolo è associato all'attivazione del pathway dell'aggrefagia, mediato dall'ubiquitinazione degli aggregati e deputato all'eliminazione degli stessi tramite autofagia.
Questa nuova patologia è quindi una malattia da accumulo di aggregati, causata da una espansione di una regione della proteina con apparente incapacità dell'aggrefagia stessa di controllarne l'accumulo.
Lo studio del meccanismo molecolare a livello cellulare, non è stato possibile nei mioblasti derivati da biopsie dei pazienti, in quanto questa proteina non risulta essere espressa a questo livello.
Pertanto, abbiamo deciso di generare un modello di overespressione nel quale la proteina mutata e quella wild-type verranno confrontate per validare la propensità della variante mutata a causare precipitazione di aggregati.

Inoltre, speriamo che questo modello possa essere utilizzato come sistema di screening per l'identificazione di composti che possano agire sulla formazione o sulla rimozione degli aggregati stessi.

1. Introduction

1.1 Distal myopathies

Distal myopathies are a heterogeneous group of disorders that have as common feature the early involvement of the distal muscles, notably hand and feet or both.

The term distal myopathy was first used by Gowers in 1902 in a case report [Gowers, 1902], even though cases describing distal muscular atrophy were earlier detailed by Roth between 1885 and 1893.

Moreover, in 1951 Welander illustrated 249 patients from 72 Swedish families affected by weakness of the long extensors of the hands [Welander, 1951].

Due to the variability of clinical presentation, the wide genetic involvement and the presence of complex multisystem involvement, a thorough evaluation of the patient with distal myopathy is of chief importance.

Some specific tests and observations might be helpful for an effective differential diagnosis.

Specifically, as reported in a recent revision of distal myopathies [Felice, 2020], some clues might come from the electromyography, the quantification of creatine kinase which is usually normal or mildly elevated in most of the distal myopathies, as well as from the family history and the pattern of muscles involved.

Moreover, the muscle imaging (MRI) might further elucidate the specific pattern of muscle deterioration thus helping in the diagnostic process.

The analysis of patients' muscle biopsy has also a crucial role in the differential analysis, since different distal myopathies might show presence of specific changes such as dystrophic and inflammatory components in Miyoshi myopathy (with mutations in the *DYSF* gene), rimmed vacuoles in myofibrillar myopathies or fiber type disproportion in Laing myopathy (*MYH7* related).

The immunohistochemistry analysis of muscle biopsies might also highlight characteristic elements such as the protein accumulation in myofibrillar myopathies.

Since the discovery of *DYSF* as the first gene related to a distal myopathy in 1998 by Brown's group [Liu et al, 1998], many more causative genes were identified with an impressive boost given in the last years thanks to the introduction of Next Generation Sequencing techniques.

Likewise, as a result of the use of NGS, it is now evident that many of these genes are liable for different clinical phenotypes.

A comprehensive review of the currently known distal myopathies from Savarese and colleagues [Savarese et al, 2020], illustrates the 19 genes so far identified and provides a practical classification of these disorders (Table 1).

Clinical entity	Gene(s)	Trasmission	References
Adult – late onset distal myopathies			
Welander distal myopathy	*TIA1*	AD	Hackman et al., 2012
Digenic SQSTM1 and TIA1 mediated distal myopathy	*SQSTM1+TIA1*	DG	Lee et al., 2018
Tibial muscular dystrophy (Udd myopathy)	*TTN*	AD	Hackman et al., 2002
Vocal cord and pharyngeal distal myopathy	*MATR3*	AD	Senderek et al., 2009
Distal Actininopathy	*ACTN2*	AD	Savarese et al., 2019
Distal Myopathy with sarcoplasmic bodies	*MB*	AD	Olive et al., 2019
Oculopharyngeal distal myopathy	*NOTCH2NLC, LRP12 and GIPC1*	AD	Deng et al., 2020; Ishiura et al., 2019; Saito et al., 2020; Sone et al., 2019
PLIN4 mutated distal myopathy	*PLIN4*	AD	Ruggieri et al. 2020
VCP distal myopathy	*VCP*	AD	Palmio et al 2011
Myofibrillar distal myopathies			
Distal myopathy with myotilin defect	*MYOT*	AD	Penisson-Besnier et al., 2006
Late onset distal myopathy (Markesbery-Griggs, Zaspopathy)	*LDB3*	AD	Griggs et al., 2007
Desminopathy	*DES*	AD > AR	Sjoberg et al., 1999
Alpha-B crystallin-mutated distal myopathy	*CRYAB*	AD	Reichlich et al. 2010
Early adult onset distal myopathies			
Miyoshi myopathy	*DYSF*	AR	Liu et al., 1998
Recessive distal titinopathy	*TTN*	AR	Evila et al., 2017
Distal myopathy with rimmed vacuoles (Nonaka and GNE myopathy)	*GNE*	AR	Kayashima et al., 2002
Distal ABD-filaminopathy	*FLNC*	AD	Duff et al., 2011
DNAJB6 distal myopathy	*DNAJB6*	AD	Ruggieri et al., 2015 - Palmio et al., 2020
Rimmed vacuolar neuromyopathy	*HSPB8*	AD	Ghaoui et al., 2016
ANO5 distal muscular dystrophy	*ANO5*	AR	Bolduc et al., 2010
RYR1 mutated calf predominant distal myopathy	*RYR1*	AD/AR	Laughlin et al., 2017 - Jokela et al., 2019
Early-childhood onset distal myopathies			
Early onset distal myopathy (Laing)	*MYH7*	AD > AR	Meredith et al., 2004
Early onset distal myopathies with nebulin defect	*NEB*	AR > AD	Wallgren-Pettersson et al., 2007, Kiiski et al., 2019
Early onset ADSSL distal myopathy	*ADSSL*	AR	Park et al., 2016
Early onset distal myopathy with KLHL9 mutations	*KLHL9*	AD	Cirak et al., 2010
Other myopathies and dystrophies with distal weakness			
Distal myopathy with caveolin defect	*CAV3*	AD	Tateyama et al., 2002
DNM2 related distal myopathy	*DNM2*	AD	Bitoun et al., 2005

Table 1. List of distal myopathies with the causative genes. (AD: autosomal dominant, AR: autosomal recessive, DG: digenic) (Savarese et al, 2020).

Hereafter, a brief overview highlighting some of the identified distal myopathies as classified in Savarese et al.

Distal myopathies with late-adulthood onset

One of the first identified entity of this group is the *Welander distal myopathy (WDM)* [Welander, 1951] defined by initial involvement of wrist and finger extensor and caused by mutations in the *TIA1* gene that encodes an RNA-binding protein that regulates the splicing of the Fas receptor and nucleates stress granules in the cell cytoplasm under stress conditions.

Moreover, described by Udd's group, is the *Tibial muscular dystrophy* known as *Udd myopathy* [Udd et al, 1993]. This myopathy, with an onset in the third or fourth decade of life and characterized by involvement of ankle dorsiflexor and atrophy of anterior lower leg muscles, is caused by mutations in the *TTN* gene, coding for the giant protein titin, part of the sarcomere. Both recessive and dominant mutations have been identified causing different clinical phenotypes, so that diseases caused by mutations in the *TTN* gene are collectively called titinopathies.

In addition, the *distal actininopathy* is caused by mutations in the *ACTN2* gene coding for alpha-actinin-2, a scaffold protein of the Z-disk interacting with titin and myotilin. Several causative variants of *ACTN2* have been associated with different phenotypes such as congenital myopathy and cardiomyopathy. The distal adult phenotype (autosomal dominant adult-onset distal myopathy-6, MPD6) is characterized by foot drop, later progressing to proximal lower limbs weakness [Savarese et al, 2019].

Noteworthy is the *distal myopathy VCP-related*, caused by mutations in the *VCP* gene encoding for the valosin-containing protein AAA+ (ATPases associated with various cellular activities) with multiple functions including organelle biogenesis and ubiquitin-dependent protein degradation. The adult distal myopathy phenotype is mostly defined by distal leg muscle weakness and foot drop with rapidly progressing frontotemporal dementia and Paget disease. The muscle biopsy shows myopathic changes and rimmed vacuoles.

Distal myopathies with myofibrillar disorganization

An early reported type of myofibrillar myopathy is *desminopathy* (MFM1), a distal myopathy originally described in 1943 by Milhorat and Wolff [Milhorat and Wolff, 1943] whose genetic cause was only identified in 1998. The *DES* gene, codes for a type III intermediate filament that connects the Z-disk with the sarcolemma and regulates the architecture of the sarcomere. The clinical

presentation begins with weakness of the hands and lower legs with a subsequent involvement of the proximal muscles. Cardiomyopathy has also been frequently reported and creatine kinase (CK) is usually slightly elevated. The MRI is suggestive, showing the initial involvement of peroneal muscles followed by tibialis anterior, gastrocnemius and soleus muscles [Fisher et al, 2008]. The muscle biopsy presents with the typical myofibrillar disorganization.

Another protein that if mutated causes myofibrillar alteration in skeletal muscles is alpha-beta crystallin, a small heat shock protein that acts as molecular chaperone interacting with desmin in the assembly of intermediate filaments. Dominant and recessive mutations have been identified in the *CRYAB* gene (also called *HSPB5*). Typical presentation of this *Alpha-B crystallin-related myofibrillar myopathy* (MFM2) is the progressive distal weakness with wasting of the lower and upper limbs, often associated with dysphagia, dysphonia and ventilatory insufficiency.

Counted in this group is also the *distal myopatilinopathy* (MFM3) caused by mutations in the *MYOT* gene, encoding for a sarcomeric protein directly interacting with alpha-actinin. The phenotype is evident after the fifth decade of life as impairment of ankle dorsiflexion and/or calf muscles. Allelic disorder is the Limb-girdle muscular dystrophy 1A (LGMD1A) that presents with overlapping clinical features especially in the later stage of the disease. The histopathological analysis of muscle biopsies shows the myofibrillar disorganization with myotilin accumulation.

The *Markesbery-Griggs distal myopathy* (MFM4), also known as *Zaspopathy*, is caused by dominant mutations in the *LDB3* gene (also called *ZASP*) and it is phenotypically characterized by initial ankle weakness progressing to proximal muscles. Similarly, the muscle biopsy shows disorganization of the myofibrils and presence of vacuoles. The product of *LDB3*, the lim domain-binding 3 protein, also interacts with other Z-disk proteins.

Distal myopathies with early-adulthood onset

One of the first described distal myopathy with young adulthood onset is *Miyoshy myopathy*, caused by recessive mutations in the *DYSF* gene coding for a protein highly expressed in skeletal muscle and involved in calcium-mediated membrane repair. Patients with mutations in this gene present with weakness of distal muscles of both upper and lower limbs, particularly affecting gastrocnemius and soleus, progressively involving thigh and gluteal muscles. Their walk is difficult but patients remain ambulant even in later stages. Muscle biopsies reveal myopathic and dystrophic changes with necrosis [Miyoshi et al, 1986]. Miyoshy myopathy is allelic to Muscular dystrophy, limb-girdle, autosomal recessive 2 (LGMDR2) Dysferlin-related, the later characterized

by initial proximal involvement yet merging in a unique dysferlinopathy at later stages of the disorders.

Another entity in this group is the recessive *GNE-related distal myopathy* with early adulthood onset also known as *Nonaka distal myopathy* [Nonaka et al, 1981]. Affected individuals show weakness and atrophy of distal muscles including hamstring and tibialis anterior, while quadriceps are spared. Electromyography (EMG) indicates myopathic changes and muscle biopsies reveal presence of rimmed vacuoles positive to acid phosphatase, indicating increased autophagic activity, and deposits of congophilic amyloid material. Mutations in the *GNE* gene, encoding for the UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase involved in sialylation, have been nowadays described in patients of many populations.

An additional member of this group of myopathies is the *distal myopathy DNAJB6-related,* in which mutations affect the N-terminal J domain of *DNAJB6* gene [Palmio et al, 2020], encoding a member of the family of co-chaperones DNAJ/HSP40. Allelic disorder is the muscular dystrophy, limb-girdle, autosomal dominant 1 (LGMDD1) in which, however, mutations affect the G/F domain of the protein.

Distal myopathies with early-childhood onset

Characterized by onset in early childhood, even though presenting with some onset variability, is *Laing myopathy*, caused by mutations in the *MYH7* gene. This gene encodes for the myosin heavy chain beta (MHC-β) isoform, expressed mainly in cardiac muscle as well as in skeletal muscles. The first symptoms are weakness of ankle dorsiflexors and toe extensors with a slow progression of the disease and CK slightly elevated, but more severe phenotypes have been reported with proximal, neck and facial muscles impairment. Moreover, there seems to be a direct correlation between phenotype and location of the mutations. In fact, mutations in the N-terminal region of the gene are mainly associated with hypertrophic cardiomyopathy, while variants in the C-terminal portion lead to hyaline body myopathy without cardiac involvement.

Likewise, an early-childhood distal myopathy is the *NEB-related myopathy*, caused by homozygous or compound heterozygous mutations in the *NEB* gene (primarily nonsense, out-of-frame indels or copy number variants, and splicing variants) and characterized by a very slowly progressive myopathy with defective extensor muscles of feet and hands. Muscle biopsy shows the presence of aggregates named nemaline bodies or nemaline rods that are composed of Z-disk and thin filaments. Nebulin is a giant actin-binding protein that regulates the assembly of the sarcomere,

the mechanical properties of the thin filament and provides stability because of its interaction with structural proteins within the Z-disk.

1.1.1 Clinical features of a novel distal vacuolar myopathy affecting an Italian family

A large Italian family is affected by distal vacuolar myopathy with late onset and slow progression. The transmission within the family shows a dominant pattern.

Patients that underwent a neurological examination, presented with a broad spectrum of severity, ranging from asymptomatic to severely affected.

The first signs of the disease are weakness of the tibialis anterior and peroneus, followed by involvement of the fingers and wrist extensors.

The most severely affected patients showed remarkable weakness of the bilateral foot dorsiflexor and neck flexor, with a less severe involvement of the shoulders and fingers' muscles, without fasciculation.

The less severe cases would show the same symptoms but delayed in time, though following the same trend.

Dysphagia and dystonia were never reported, at any time in patients' life. Moreover, cardiac muscle seems to be preserved in its functionality, as well as the respiratory functions.

Electromyography (EMG) showed myopathic potentials in proximal and distal limb muscles in all affected individuals and frequent repetitive high frequency discharges.

Creatine kinase (CK) levels were doubled in the most severely affected patients compared to the controls, while in the least severe, the CK levels were within the normal range.

Microsatellite and SNP array analysis revealed that this disease is not related to any other distal myopathy so far reported, indicating as common haplotype in a region of 5.12 Mb in the short arm of chromosome 19 (19p13.3).

1.2 Lipid Droplets

Despite having been long considered as mere cytoplasmic lipid deposits, it is nowadays universally acknowledged that Lipid Droplets (LDs) are dynamic organelles, able to move within the cell cytoplasm and to interact with various organelles.

LDs are present in virtually all cell types in all eukaryotes, and all cell types have the ability to generate LDs when exposed to high concentration of fatty acids.

LDs are composed of a hydrophobic core that contains neutral lipids, ringed by a monolayer of polar lipids. The surface of this structure is decorated by more than 200 proteins such as structural proteins of the perilipin family, enzymes for the synthesis of lipids such as diacylglycerol acyltransferase (DGAT), caveolins and membrane-trafficking proteins such as the members of the Rab family [Guo et al, 2009].

Depending on the cell type and on the physiological cell state, the LDs' composition of lipids and proteins varies, supporting the idea that LDs are not just a storage space.

The main function of LDs is the storage and mobilization, when needed, of fatty acids and cholesterol, such as diacylglycerols, triacylglycerols (TGA), cholesteryl esters, retinyl esters and free cholesterol in diverse proportion depending on the cell type. In particular, fatty acids are esterified to be stored progressively as mono, di and triacylglycerols. Moreover, LDs are source of lipids for membrane components and signaling molecules accumulated as cholesteryl esters to sequester excessive cellular cholesterol [Bickel et al, 2009].

Neutral lipids are physiologically essential in many cell types. For instance, cells of the adrenal cortex, ovary and testis contain many LDs enriched in cholesteryl ester that are the precursors of steroids. Moreover, LDs are fundamental storage for surfactant production in lung's epithelia cells, contain retinyl esters as precursors of vitamin A in liver stellate cells or store retinyl esters as precursors a chromophore of rhodopsin in retinal pigment epithelia [van Meer, 2001; Martin & Parton, 2005; Guo et al, 2009].

Fatty acids can either be synthesized within the cells from carbohydrates or else can enter the cells passively, facilitated by the fatty acid translocases. Then, fatty acids are conjugated with CoA and form fatty acyl-CoA (reaction that requires energy) which is in turn used by the glycerol-3-phosphateacyltransferase and sn-1-acylglycerol-3-phosphate acyltransferase to generate diacylglycerols. The latter are converted by the DGAT enzymes, into triacylglycerols which are neutral lipids or else they enter the phospholipid synthesis pathway.

Sterols instead are either synthesized or internalized into the cells through endocytosis and degradation of lipoproteins in the lysosome.

Different models could be hypothesized for the formation of LDs (Fig.1).

The canonical model, defined as *ER-budding model*, supposes that, within the ER bilayer, a neutral lipid lens is formed and then it buds from the ER membranes, taking phospholipids from the cytosolic side. A modification of this model, called the *ER domain model*, assumes that the LDs won't dissociate from the ER but will remain as specialized domain, a protrusion of the ER. [Guo et al, 2009]

Another model, theorized by Ploegh and colleagues and termed *bicelle model* [Ploegh, 2007], is similar to the ER-budding model since the lipids accumulate within the ER membrane. Another hypothesis is that the generation of the LDs is due to an excision that involves both sides of the ER membrane, the luminal and the cytosolic ones. Issues of this models are the fact that the excision would cause a temporary hole in the ER through which ions could escape, thus locally modifying the redox environment as well as the mechanism by which the LD would move toward the cytoplasm instead of going to the ER lumen.

The last model is the *vesicular-budding model* proposed by Walther and Farese [Walther & Farese, 2008]. In this model a small bilayer vesicle is produced from the cytoplasmic side of the ER, likely in specialized domains and remains attached to it. The vesicle will then be filled with neutral lipids by a shuttling mechanism and the luminal portion of the membrane will be squeezed into the vesicle and remain within. This model is consistent with the observation that hydrophilic proteins have been identified inside the LDs.

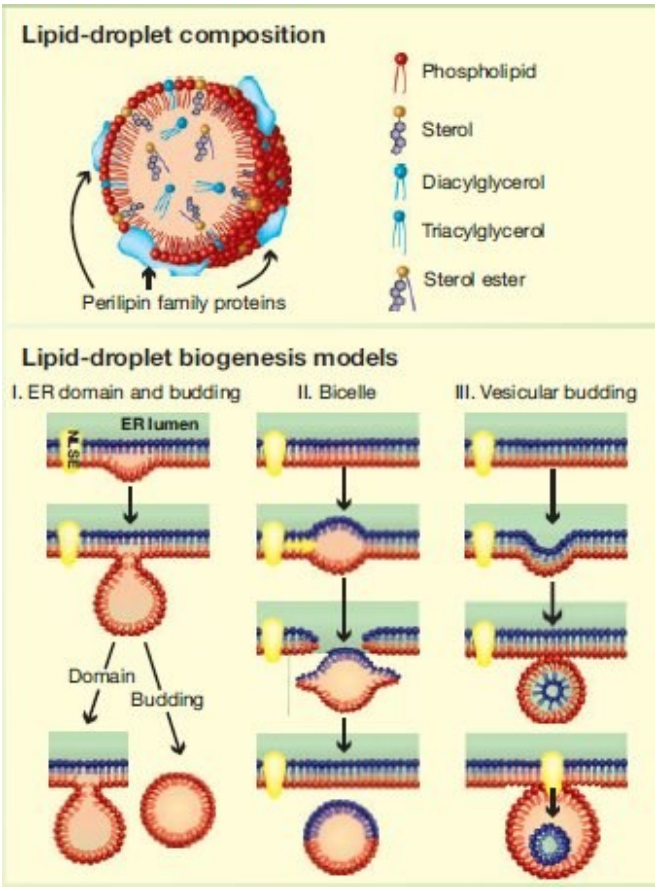


Figure 1. Representation of the LDs composition and proposed models for biogenesis (Guo et al, 2009).

After their formation, LDs are able to grow, as demonstrated by the wide range in size that can be observed in different cell types as well as within the same cell cytoplasm. For instance, white adipose cells contain a giant LD with a 100 µm diameter that occupies the entire cytoplasm, while non-adipocyte LDs are generally no more than 10 µm in diameter and are sparse within the cytoplasm but far from the cell membrane (Fig.2) [Suzuki et al, 2011].

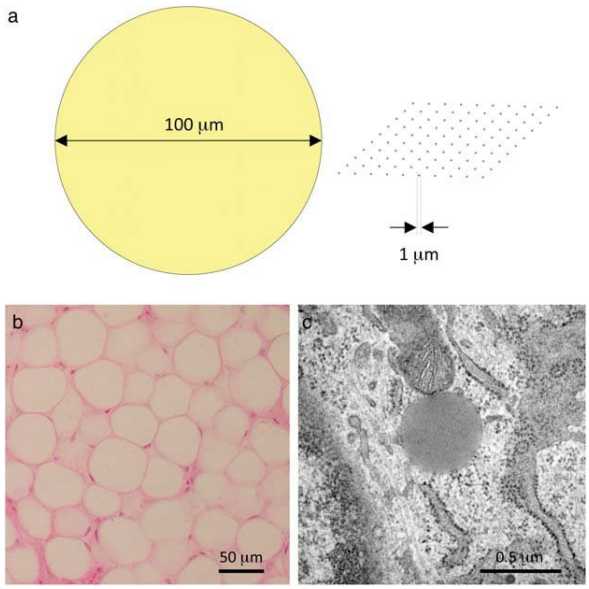


Figure 2. **a** Drawing of white adipocyte LD (100 µm diameter) versus a non-adipocyte LD (1 µm diameter). Cells of the white adipose tissue in mouse (**b** light microscopy image) with giant LDs occupying the entire cell cytoplasm compared to the smaller LDs in rat fibroblasts (**c** electron microscopy image) supplemented with oleic acid (Suzuki et al, 2011).

The mechanism by which new neutral lipids are pumped into the formed LDs is yet to be clarified. It is possible though that in the case of LDs that don't bud from the ER, lipids and protein will be able to diffuse from the ER itself, while if the LDs have detached, the delivery could happen via vesicular transport with the issue though that a bilayer membrane (the vesicular transport one) will have to fuse with a monolayer membrane (the LD membrane).

Furthermore, the neutral lipids could be synthesized locally by DGATs present on the surface of the LDs.

Or else, LDs could fuse with each other to form larger LDs [Guo et al, 2009].

LDs within the cytoplasm are able to move with directional movements for long distances (hundreds of nanometres per second) and for long time (over seconds to minutes) or they can move with short Brownian movements. The anterograde and retrograde movements are mediated by kinesin-1 and dynein proteins, respectively [Welte, 2009; Suzuki et al, 2011].

The reason for this motility is possibly to be searched in the need to supplement with lipids or to shuttle proteins among different membranous compartments within the cell [Guo et al, 2009; Welte, 2009].

In skeletal muscle, a tissue with a high energy requirement, numerous LDs are localized near the sarcolemma or amidst the myofibrils (Fig.3). Their size ranges from 0.3 µm up to 1.5 µm with LDs in type 1 fibers being larger than in type 2.

Being localized within the myofibrils, the LDs are subject to passive movements during muscle contraction but can also be conveyed by microtubules and intermediate filaments.

The three lipases involved in myocellular lipolysis are the monoacylglycerol lipase (MAG), the adipose triglyceride lipase (ATGL) and the hormone sensitive lipase (HSL).

The uptake of fatty acid in the skeletal muscle is possible through the binding with proteins and transporters. Once internalized, the FA can either be immediately used for energy production or else transformed into TAG and stored into LDs [Li et al, 2014; Bosma, 2016; Morales et al, 2017].

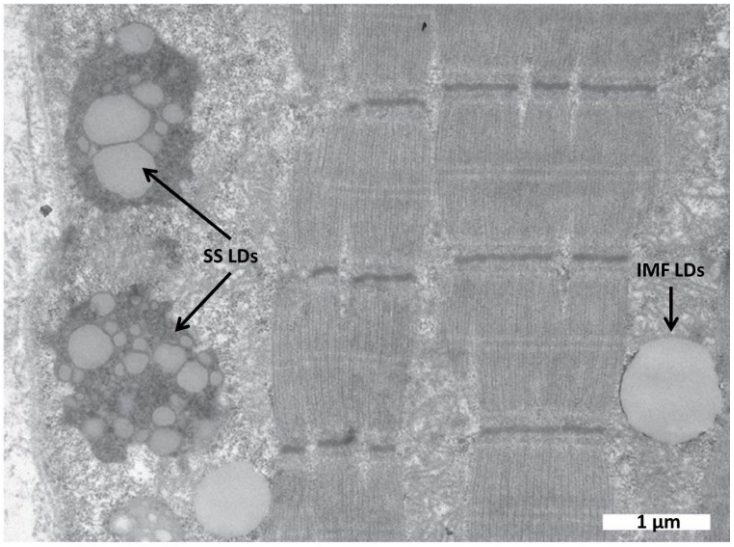


Figure 3. An electron microscopy image of a section of skeletal muscle, showing subsarcolemmal LDs (SS LDs) and intermiofibrillary LDs (IMF LDs) (Li et al, 2014).

The lipolysis, the degradation of triacylglycerols to produce glycerol and free fatty acids (FFA), can happen only upon removal of the proteins, specifically perilipins, that surround the LDs protecting them from unwanted degradation of TGA.

When TGA need to be processed, perilipins are phosphorylated by the protein kinase A so that the first lipase, the adipose triglyceride lipase (ATG), also known as patatin-like phospholipase domain-containing protein 2 (PNPLA2), can initiate TGA hydrolysis.

The ATG action begins by transforming TGA into diacylglycerol and FFAs; its action is activated by the lipid droplet-binding protein CGI-58 and blocked by the G0/G1 switch protein 2 (G0S2).

The following step in lipolysis is accomplished by the hormone sensitive lipase (HSL), while the last step is carried out by the monoacylglycerol lipase (MGL) which cleaves both triacylglycerol-derived and glycerophospholipid-derived monoacylglycerols.
Unlike ATG and HSL, MGL is cytoplasmic, seen the higher solubility of monoacylglycerols [Guo et al, 2009; Wang, 2016; Zechner et al, 2009; Zechner et al, 2017].

Degradation of lipids can also happen within the lysosome, by the lisosomal acid lipase (LAL).
LDs, in fact, can undergo lipophagy, a specialized form of autophagy, in which small or portions of LDs are absorbed by LC3-II-positive membranes. After degradation by LAL, the FFAs are released into the cytosol [Wang, 2016; Zechner et al, 2017].
Moreover, perilipin 2 and perilipin 3 were reported to be selectively degraded by CMA during starvation [Kaushik & Cuervo, 2015].

1.3 Perilipins

Perilipins are members of a family of proteins that associate to LDs regulating lipolysis.
The family is composed of 5 members, whose nomenclature has been revised and unified in 2010, now defining *PLIN1* (previously called Perilipin), *PLIN2* (also known as ADRP, adipose differentiation-related protein), *PLIN3* (TIP47, tail-interacting protein of 47 kiloDaltons), *PLIN4* (S3-12) and *PLIN5* (OXPAT, Oxpatperilipin) [Kimmel et al, 2010].
Perilipins, even though presenting with a certain degree of structural similarity, differ from each other for tissue and cellular localization, size, affinity for lipids and stability when not bound to lipids.
Particularly, perilipin 1, 4 and 5 are tissue specific, while perilipin 2 and 3 are ubiquitous. Moreover, perilipin 1 and 2 are unstable and rapidly degraded when not bound to LDs, while the other members of the family are stable and have exchangeable binding properties.
Furthermore, some of the perilipins preferentially associate with LDs rich in triacylglycerol (perilipin 2, 3 and 5), while perilipin 1 and 4 favor LDs enriched in cholesterol ester [Hsieh et al, 2012].
Evolutionarily, perilipins seem to have originated from genome duplications from an ancestral precursor to form perilipin 1 and a precursor of perilipin 2, 3, 4 and 5. During a second genome duplication there was a further partition that divided perilipin 2 from the precursor of perilipin 3, 4 and 5. More recently, another duplication gave rise to perilipin 4 and 5. Interestingly, zebrafish

possess a gene encoding for perilipin 6 that is only found in fish and that appears to have originated from the precursor of perilipin 1 (Fig.4) [Granneman et al, 2017].

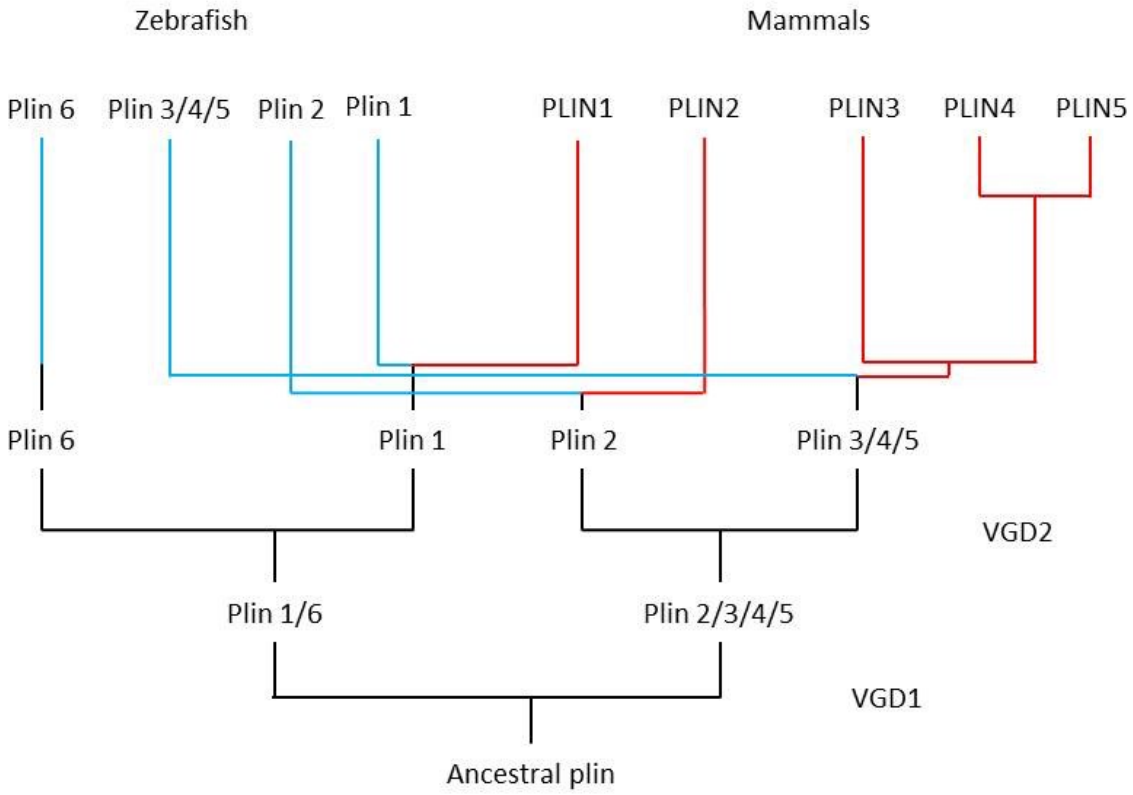


Figure 4. Diagram of the duplication events that originated the family of perilipins (modified from Granneman et al, 2017).

All perilipins but perilipin 4, possess a ~100 amino acids N terminal PAT domain whose name was given after the three perilipin A, ADRP, and TIP47 that were the first identified to interact with LDs. Moreover, in the C-terminal region of perilipin 3 a four-helix bundle was identified. This region is also present in perilipin 1, 2, 3 and 5 but not in perilipin 4.

In addition, another shared domain is the amphipathic helix (AH), responsible for the binding with LDs, constituted by 11-mer imperfect repeats (Fig.5). These helices are called 3-11 because each turn contains 3.67 amino acids and to obtain an integer, 3 turns are necessary [Giménez-Andrés et al, 2018].

This kind of α-helix is also present in apolipoproteins as well as α-synuclein [Segrest et al, 1992; Bussell & Eliezer, 2003].

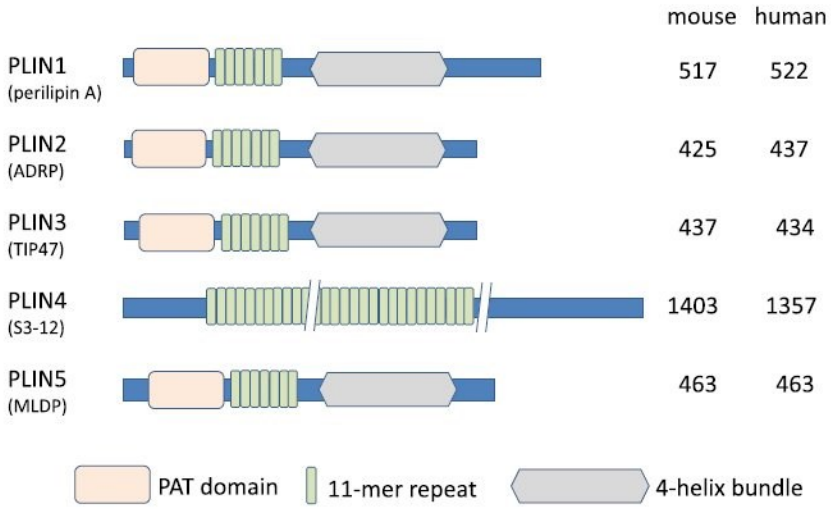


Figure 5. Schematic representation of the perilipin family members. Neither the PAT domain nor the four-helix bundle are present in perilipin 4 which is the more divergent member of the family. All of them though possess the 11-mer repeat in variable number (Itabe et al, 2007).

Perilipin 1

Perilipin 1 is the most studied protein of the family and in human is coded by the gene *PLIN1* on chromosome 15, producing three isoforms by alternative splicing, perilipin A, B and C, with perilipin A being the most abundant isoform in adipocytes LDs. These isoforms share the N-terminal region with different C-termini.

Perilipin 1 is expressed mainly in white and brown tissue adipocytes and to a lesser extent in steroidogenic cells of the adrenal cortex, testis and ovaries. Moreover, it is stable only if associated to lipid droplets, with a half-life of 40 hours, while it is rapidly degraded if free in the cytoplasm.

Perilipin 1 possess consensus sites for phosphorylation that appear to be essential for the initial step of lipolysis. Under basal conditions both perilipin 1 and ATGL co-localize on the LD surface and perilipin 1 interacts with CGI-58, a co-activator of adipose triglyceride lipase. When neutral lipids need to be mobilized, perilipin 1 is phosphorylated in a c-AMP dependent manner which causes the release of CGI-58 that can in turn activate ATGL [Brasaemle, 2007; Bickel et al, 2009; Itabe et al, 2017; Sztalryd & Brasaemle, 2017].

The knock out murine model of perilipin 1 is lean, with very small LDs in the adipose tissue and has higher glucose tolerance and resistance to obesity due to the diet [Martinez-Botas et al, 2000; Tansey et al, 2001]. Frameshift mutations in *PLIN1* gene have been reported to cause lipodystrophy type 4 (MIM # 613877), a dominant inherited disorder due to haploinsufficiency.

The disorder, with childhood or young adult onset, presents with loss of subcutaneous adipose tissue, diabetes mellitus with insulin resistance, hypertriglyceridemia, and hypertension [Gandotra et al, 2011].

Perilipin 2

Perilipin 2 is a 50 kDa protein coded by the gene *PLIN2* on chromosome 9, that is highly expressed in adipose tissue as well as in skeletal muscle. Moreover, it is the most expressed LD protein in hepatocytes.

Similarly to perilipin 1, it is unstable if not bound to LDs and it is rapidly degraded via the ubiquitin/proteasome pathway [Xu et al, 2005].

In adipose tissue, perilipin 2 contributes to differentiation and regulation of LDs and it is present in premature LDs but it is replaced by perilipin 1 in more mature LDs.

Despite regulating LDs, it does not recruit lipase to their surface under hormonal stimulation, unlike perilipin 1 and 5 [Bickel et al, 2009; Morales et al, 2017; Itabe et al, 2017; Sztalryd & Brasaemle, 2017].

Perilipin 3

Perilipin 3 (also known as tail-interacting protein of 47 kDa, TIP47) is coded by the *PLIN3* gene located on chromosome 19. It was initially implicated in cytosolic transport, particularly for its binding to the cytosolic domain of the mannose-6-phosphate receptor (M6PR) which will cause its inclusion into vesicles.

The crystal structure of the four-helix bundle of perilipin 3 was clarified by Hickenbottom and colleagues [Hickenbottom et al, 2004]. This structure is similarly present in ApoE and ApoA1 which are able to detach and move between lipoprotein particles via such α-helices. Moreover, in perilipin 3, a gap is formed by the four-helix bundle, probably responsible for recruitment and binding of LDs [Ohsaki et al, 2006]. By homology, similar structures are also present in perilipin 1, 2 and 5.

Perilipin 3 along with perilipin 4 is the first to coat LDs, then perilipin 2 and perilipin 1 intervene.

The exact function and involvement of perilipin 3 in lipolysis have yet to be clarified [Bickel et al, 2009; Morales et al, 2017; Itabe et al, 2017; Sztalryd & Brasaemle, 2017].

Perilipin 5

As hinted by the original name OXPAT given by Wolins and colleagues [Wolins et al, 2006], perilipin 5 is the most expressed LD protein in oxidative tissues, such as heart, slow-twitch muscle brown adipose tissue, and liver.

It is coded by *PLIN5* gene on chromosome 19 and its amino acid sequence has a certain degree of identity with perilipin 3 (30%) and perilipin 2 (26%) and like perilipin 3 and 4 it is stable even when not tied to LDs but free in the cytosol. It is immediately summoned to LDs when conditions promote their formation.

In heart, liver and skeletal muscle, perilipin 5's transcription is regulated by peroxisome proliferator-activated receptor α (PPARα) and PPARδ [Yamaguchi et al, 2006; Wolins et al, 2006; Bindesbøll et al, 2013]. It has been shown that perilipin 5 in cardiac and skeletal muscle localizes to mitochondria, suggesting that perilipin 5 might facilitate the transfer of free fatty acids to the mitochondria during lipolysis to help mitochondria oxidation [Bosma et al, 2012].

1.3.1 Perilipin 4

In humans perilipin 4 is coded by the *PLIN4* gene which is located in tandem with the *PLIN5* gene on chromosome 19. It has the more diverse sequence compared to the other perilipins, lacking the PAT domain and the four-helix bundle, suggesting that it might have evolved for specific functions. Perilipin 4 sequence is also peculiar, having the longest amphipathic helix (AH) spanning almost 1000 amino acids with a unique repetitiveness.

Its expression is mainly evident in white adipose tissue, with detectable positivity also in brown adipose tissue, heart and skeletal muscle.

Wolins and colleague [Wolins et al, 2003] demonstrated that upon stimulation of TGA synthesis in the adipocytes, the freshly formed LDs were coated with perilipin 4, suggesting that this protein is involved in the early formation of lipid droplets. Later on the same group, showed that a change in LDs protein composition is visible after some time from oleic acid treatment, going from the initial presence of perilipin 2, 3 and 4 on small LDs, to a situation in which larger LDs are coated with perilipin 2, while the small peripheral LDs are still mainly coated with perilipin 3 and 4 (Fig.6) [Wolins et al, 2005].

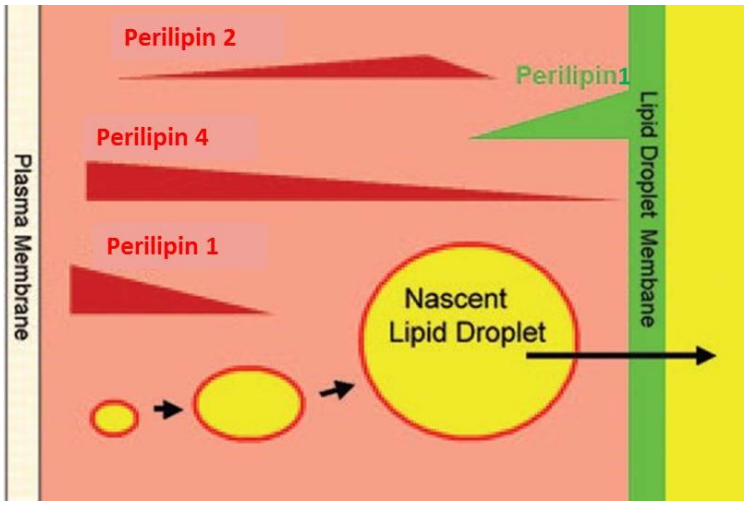


Figure 6. Model for perilipins involvement in the biogenesis of lipid droplets (Wolins et al, 2005).

In skeletal muscle, LDs are mainly localized in the subsarcolemmal (SS) region of the fibers as well as in the intermyofibrillar region (IMF) [Malenfant et al, 2001]. The mRNA of perilipin 4 is the highest expressed compared to the other perilipins but it was demonstrated that after 12 weeks of strength and endurance exercises, its quantity was reduced in vastus lateralis and that the major perilipin 4 protein staining was present in the SS LDs. Moreover, perilipin 4 staining was primarily localized at the subsarcolemmal region of the fibers, mainly in type I fibers that have an oxidative metabolism and contain more neutral lipids, that in type II fibers (Fig.7) [Li et al, 2014; Pourteymour et al, 2015].

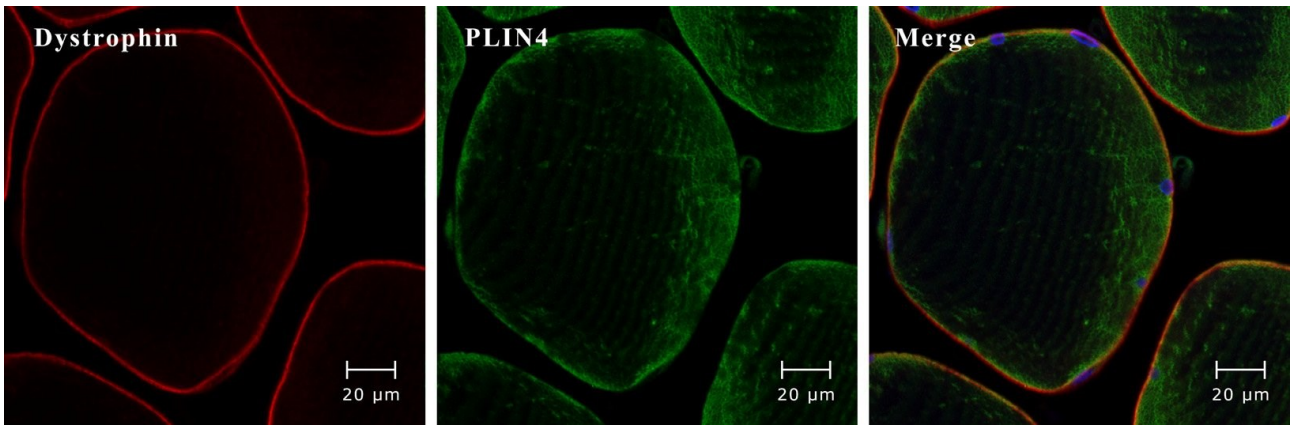


Figure 7. Distribution of perilipin 4 in skeletal muscle fibers. The subsarcolemmal localization is evident when compared to the dystrophin staining (Pourteymour et al, 2015).

The amphipathic helix (AH) of perilipin 4 was only recently extensively studied by Čopič's group [Čopič et al, 2018].

Compared to the other perilipins, in perilipin 4 the 11-mer repeats of the AH, are well conserved at the level of 33-mer (3x11-mer) and are present in number of 29 tandem repeats, producing a very long region.

Similarly to perilipin 3, 5 and other lipid exchange proteins such as α-synuclein and apolipoproteins, perilipin 4 is intrinsically disordered when free in the cytoplasm, folding into the 3-11 helix only when contacting an appropriate interface.

Its amino acidic composition is peculiar, being enriched in some amino acids like threonine, glycine, and valine and almost totally lacking aromatic residues, thus giving low hydrophobicity to this AH. Seen perilipin 4 AH composition, it is most likely that its binding to LDs is possible because of the many weak interactions distributed along the entire length of the AH (Fig.8) [Čopič et al, 2018].

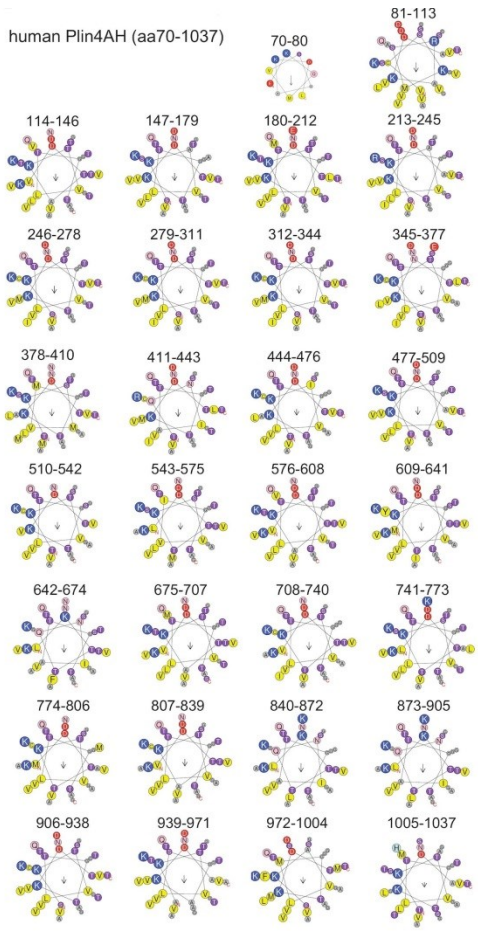


Figure 8. Amino acid sequence of the human perilipin 4 AH, showing the repeats organized in 33-mers. Of these imperfect repeats, 3 of them (aa 246-278, 279-311 and 312-344) are exactly identical (Čopič et al, 2018).

Moreover, Čopič and colleagues produced different peptides of perilipin 4 AH with different number of repeats, expressing them in HeLa cells supplemented with oleic acid. They

demonstrated that by increasing the number of 33-mer the positivity for perilipin 4 on formed LDs increased. For instance, the transfected 2x 33-mer was completely cytosolic while the 4x, 8x or 20x 33-mer were more and more localizing to LDs.
In a subsequent study the same group showed that the interaction between perilipin 4 and LDs is very stable and lasting for many days, unlike for other perilipins that have a more dynamic exchange with LDs [Giménez-Andrés et al, 2021]

1.4 Protein misfolding and aggregation

Proper folding of newly synthesized proteins is an exigent process especially in crowded environments such the cells. In vivo, the folding of a protein synthesized on ribosomes, might be co-translational, might happen in the cytoplasm after been released from the ribosomes or it can occur in specific compartments such the endoplasmic reticulum (ER).
Protein folding is driven by kinetic and thermodynamic factors that lead to the formation of a polypeptide with the most stable conformation [Dobson, 2003; Markossian & Kurganov, 2004].
Failure to fold correctly, might be caused by a variety of reasons such as mutations in the gene sequence that modify the amino acid sequence, the concentration of the protein, errors in the transcription process or translation, as well as oxidative or thermal stress and toxic chemicals and also the absence of molecular chaperones [Fink, 1998].

To control and correct potential misfolding of the proteins, a Protein Quality Control (PQC) system exists within the cells. Components of these PQC are the molecular chaperones, called heat shock proteins (HSPs) and classified based on their sizes Hsp40, Hsp60, Hsp70, Hsp90 and small HSPs.
Chaperones, which account for the 10% of the total proteome in the cell, recognize the hydrophobic sequences that are usually hidden in the native folding and promote refolding of proteins in three different ways. First of all, some chaperones facilitate spontaneous folding by temporarily keeping a protein in its unfolded state till its proper conformation is spontaneously reached. Second, some chaperones are capable to unfold misfolded protein modifying them into re-foldable species, by using ATP. Third, some chaperones can act on already assembled aggregates, unfolding and solubilizing the aggregates into refolded proteins, again using the energy of ATP hydrolysis [Ciechanover & Kwon, 2017].
In misfolded proteins, hydrophobic areas that are normally hidden within the native proteins, are exposed to the surface, leading to several regions that can interact in aggregates formation.

1.5 The ubiquitin-proteasome pathway (UPP)

Within the cell environment, the proteome undergoes constant synthesis and degradation during which the peptide bonds are hydrolyzed, releasing free amino acids as well as the energy that was gathered during synthesis.
Moreover, this mechanism of degradation is fundamental for the cell as a quality control system to remove damaged or misfolded proteins.

One of the major pathways for protein degradation is the Ubiquitin-Proteasome pathway (UPP).
The degradation via the UPP involves two distinct and sequential steps: first the conjugation of several ubiquitin molecules to the protein to be removed, second the degradation of such marked proteins by the 26S proteasome.

1.5.1 Ubiquitination

Ubiquitin is the most studied member of a family of proteins with diverse amino acid sequences but similar folding structure. It is a small protein of 76 amino acids highly conserved through the evolution, found in all eukaryotes but not in the prokaryotes. Its conservation is striking in human and yeast, with only 3 different amino acids out of the 76.
Ubiquitin possesses 7 lysine residues that can be ubiquitinated producing isopeptide chains. Moreover, the first methionine could also be a site of ubiquitination (Fig.9). Predominant site of protein ubiquitination in cells is the Lys48, a signal for degradation through the proteasome, while the second most abundant site, Lys68 is often related to non-proteolytic roles in subnuclear trafficking, DNA damage tolerance as well as other processes [Pickart & Eddins, 2004; Swatek & Komander, 2006; Clague & Urbé, 2010].

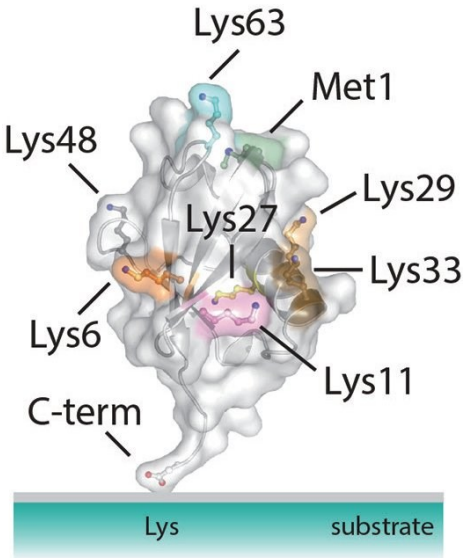


Figure 9. Structure of ubiquitin with the sites of ubiquitination. (Swateck & Komander, 2006).

Conjugation of ubiquitin to a substrate, is happening in a three-step cascade, involving subsequentially three classes of enzymes: the ubiquitin-activating enzyme E1, the ubiquitin-conjugating enzyme E2, and a substrate-specific ubiquitin-protein ligase E3. Mammalian cells possess one single E1 enzyme interacting with approximately 50 distinct E2 enzymes that in turn cooperate with over 1000 E3 enzymes that are responsible for the specific recognition of the different substrates [Reinstein & Ciechanover, 2006].

Briefly, at first the E1 enzyme generates a E1-ubiquitin thiol ester intermediate, reaction that requires ATP, followed by the transfer of the activated ubiquitin from the E1 cysteine to an E2 cysteine, step that once more requires ATP. Finally, one of the E3 ligases catalyzes the covalent attachment of the ubiquitin to the substrate. There is only a finite set of substrates for each E3 ligases [Ciechanover & Iwai, 2004; Pickart & Eddins, 2004]. The recurrence of this process allows the synthesis of polyubiquitin chains attached to the protein substrate, that can thus be recognized by the proteasome 26S for degradation.

A set of other enzymes that are responsible for the efficient removal of the ubiquitins from their substrates, also exist. These enzymes, called deubiquitinating enzymes (DUBs), are cysteine proteases and are classified based on their active sites into UCHs (ubiquitin COOH-terminal hydrolyses) or UBPs (ubiquitin-specific proteases). DUBs are fundamental to maintain the cellular free ubiquitin pool, to edit poly-ubiquitin chains and also as proofreading mechanism to correct incorrect ubiquitination and [Amerik & Hochstrasser, 2004; Glickman MH, Ciechanover, 2002].

1.5.2 Proteasome

The proteasome 26S is a multicomplex of 2.5MDa composed of two subunits, a catalytic core particle (CP), known as 20S, and one or two regulatory particles (RPs) named 19S (Fig.10)

The 20S catalytic core is a barrel-shaped complex composed of four stacked heptameric rings, two α outer and two β inner subunits. These subunits are organized in a α_{1-7}, β_{1-7}, β_{1-7} and α_{1-7} structure. Three of the β subunits, possess active threonine residues at their N-terminal regions, thus defining the proteasome as a threonine protease [Tanaka, 2009; Ciechanover & Iwai, 2004; Budenholzer et al, 2017]. The passage through the inner channel of the proteasome is controlled by the N-termini of the α subunits.

The 19S regulatory particles are identified into a lid and a base. The lid is composed of subunits called Rpn (Regulatory particle of non-ATPase) that are responsible for substrate deubiquitylation while the base is made of Rpn and Rpt (Regulatory particle of triple-ATPase) which recognize and capture the substrates, participate in substrate unfolding and open the channel [Tanaka, 2009; Huang et al, 2016; Budenholzer et al, 2017].

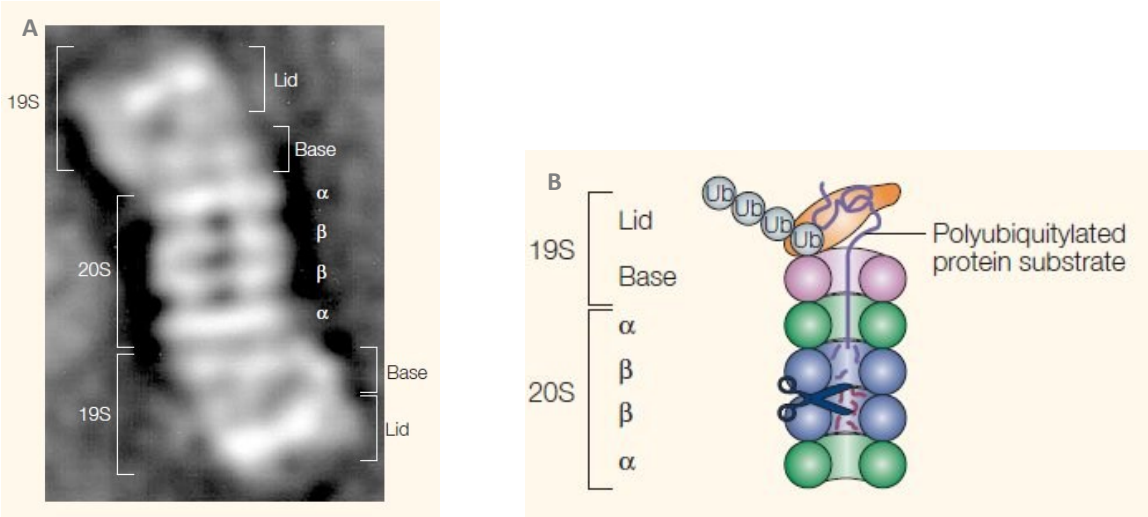


Figure 10. Electron microscopy image showing the Saccharomyces cerevisiae proteasome (A) and its simplified graphical view (B). (Ciechanover, 2005).

The substrates that had entered the proteasome are digested into small peptides from 3 to 25 residues. In the cell cytoplasm, these peptides are soon digested by other enzymes,

endopeptidases and aminopeptidases, thus producing single amino acids that can be reused by the cell.

1.6 Autophagy

The first description of a vesicle surrounded by a single or double membrane and containing fragmented organelles and portion of cytoplasm, was given in 1963 during a conference on lysosomes, by the cytologist and biochemist Christian De Duve, the recipient of the 1974 Nobel Prize in Physiology or Medicine for the discovery of such organelles back in the fifties. During this conference, De Duve conceived the term autophagy from the Greek “self-eating” to name this phenomenon.

Years later, in 2016, the Nobel Prize in Physiology or Medicine was awarded to the cell biologist Yoshinori Ohsumi for his fundamental studies on the autophagy machinery and the discovery of the Autophagy-related genes (Atg), which shed a light on a fascinating yet complex, ubiquitous cellular mechanism [Galluzzi et al, 2017].

Autophagy is a highly conserved process, existing in all eukaryotes, that allows an efficient turnover of proteins and organelles. It is in fact, along with the proteasome, one of the pathways responsible for protein degradation and the only capable of digesting entire organelles.

In mammalian cells, there are three types of autophagy, all characterized by the final lysosomal degradation. These are, chaperone-mediated autophagy (CMA), microautophagy and macroautophagy (generally indicated simply as autophagy) (Fig.11).

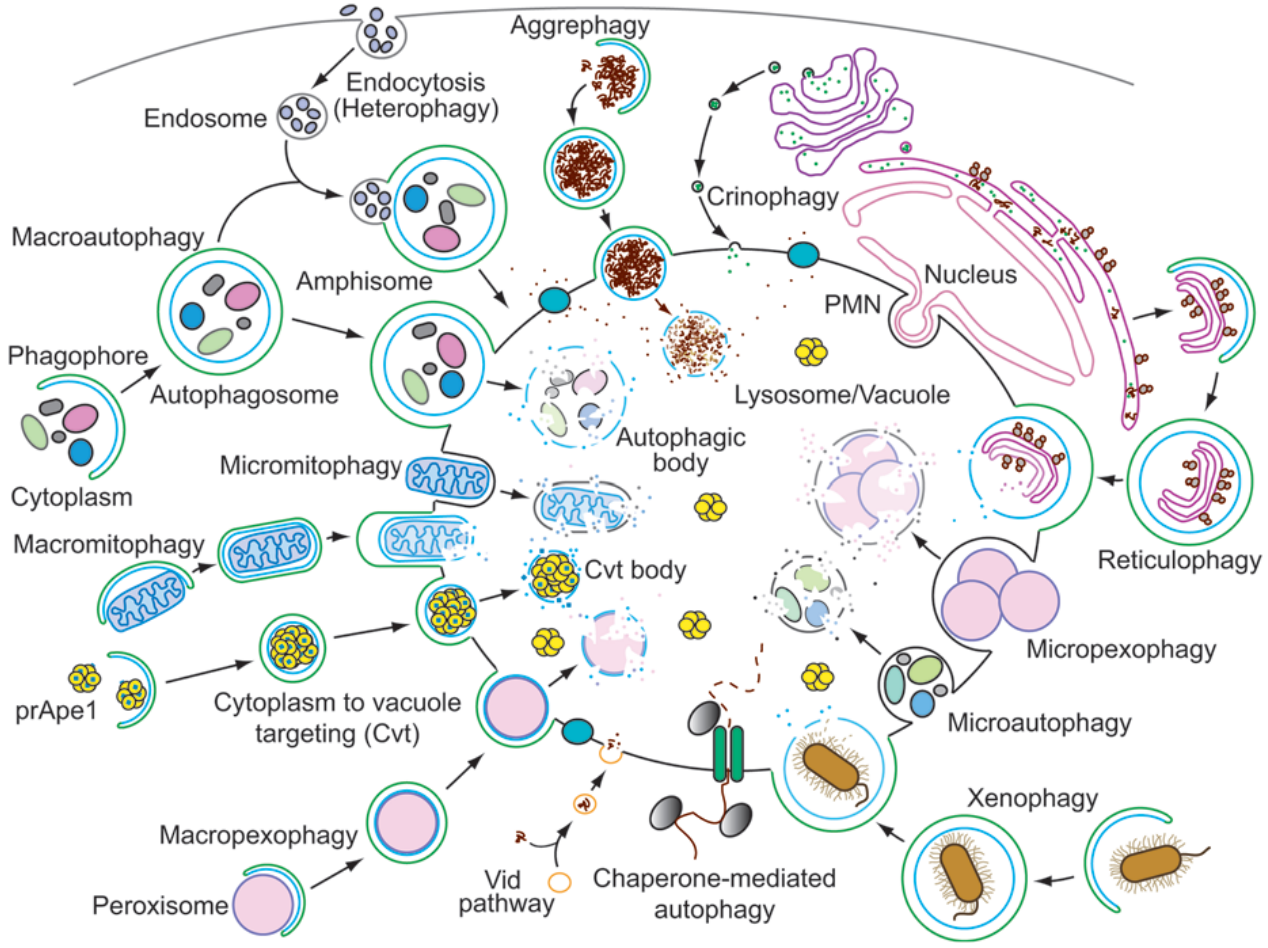


Fig.11 A depiction of the mammalian and yeast degradative pathways involving the lysosomes and vacuoles shows the complexity and variety of substrates that can be targets of these systems (2007_Klionsky).

1.6.1 Chaperone-mediated autophagy (CMA)

Chaperone-mediated autophagy has been described in higher eukaryotes and its only substrates are cytosolic proteins that are directly delivered to the lysosome for degradation, without formation of vesicles.

Dice and colleagues showed in 1990 that for proteins to be substrates for CMA degradation, they had to contain a KFERQ-like motif, an amino acid sequence that is necessary and sufficient for the binding with the cytosolic chaperone Hsc70 (also known as HSPA8) [Dice, 1990].

Nearly 40% of the total mammalian proteins contain this motif, therefore could be target for CMA degradation.

Once the substrate is bound to HSC70, is then delivered to the lysosomes, where the monomeric form of LAMP2A, another fundamental player of this CMA system, is stimulated to form an oligomeric translocation complex. The protein to be degraded is thus internalized into the lysosome for enzymatic digestion.

LAMP2A is a spliced isoform of the LAMP2 gene, along with LAMP2B and LAMP2C, which are not involved in CMA but in macroautophagy [Galluzzi et al, 2017; Kaushik & Cuervo, 2018] (Fig.12).

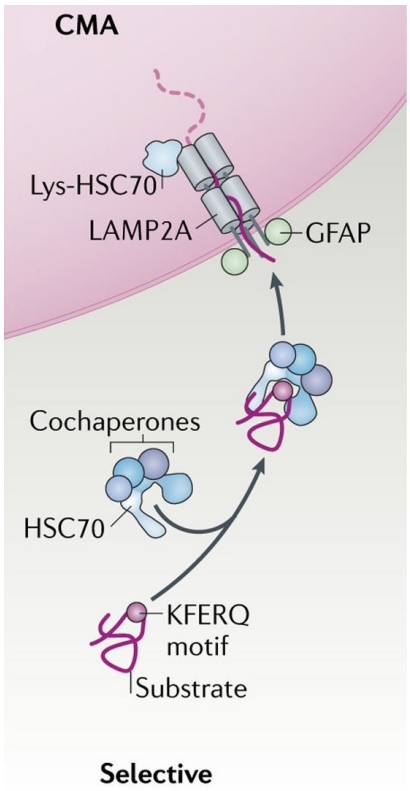


Figure 12. The process of chaperone mediated autophagy involves the recognition of the substrate via its KFERQ motif by a chaperone (Hsc70) and cochaperones, its delivery to the lysosome and internalization through the LAMP2A receptor (Kaushik & Cuervo, 2018).

1.6.2 Microautophagy

Microautophagy, is the selective or non-selective mechanism for degradation of cytoplasm, organelles and vacuole membrane proteins that does not require an autophagosome, but relies on the direct internalization in lysosomes or endosomes via small vesicles forming at the surface of these two compartments.

Lysosomes and endosomes undergo two types of morphological rearrangements: an initial invagination forming a bud with engulfment of the substrate, followed by the internal closure and detachment of this bud to form a microautophagic body.

As it happened for macroautophagy, the major understanding of the molecular machinery implicated in microautophagy, came thanks to the studies in yeast, which showed that virtually any cellular structure can be a substrate for microautophagy degradation, such peroxisomes,

cytosol, mitochondria, nuclear portions, lipid droplets, endoplasmic reticulum, some cytosolic enzymes and vacuole membrane proteins (Fig.13).

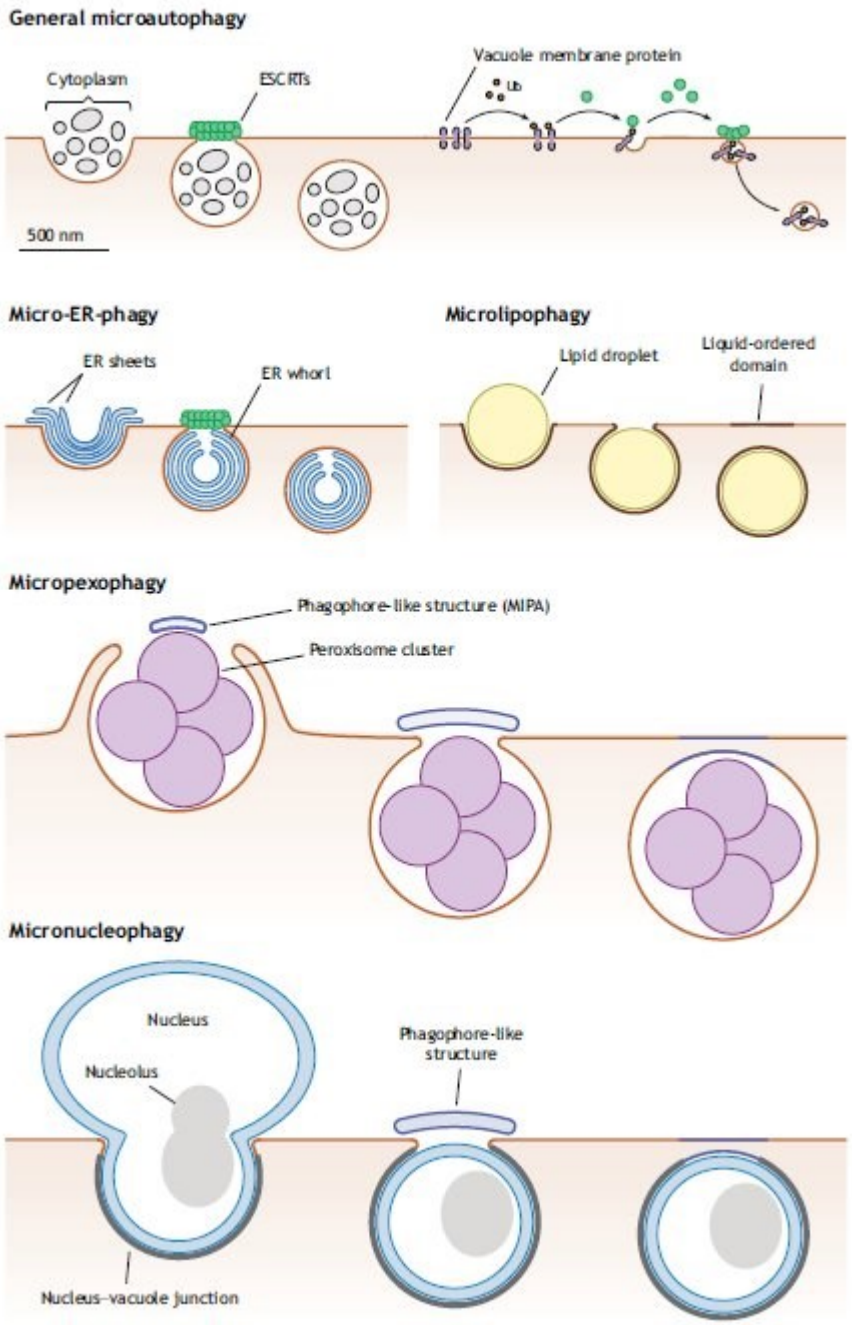


Figure 13. Microautophagy could be generalized or targeting specialized cellular compartments (Schuck, 2020).

Moreover, one of the major substrate for degradation via microautophagy are the receptors for macroautophagy, showing that these two pathways are interdependent.

Recently, a classification for the existing types of microautophagy has been proposed, based on the molecular mechanism involved in the formation of the lysosome and endosomes invaginations [Schuck, 2020].

We can then distinguish a fission-type and a fusion-type microautophagy (Fig.14).

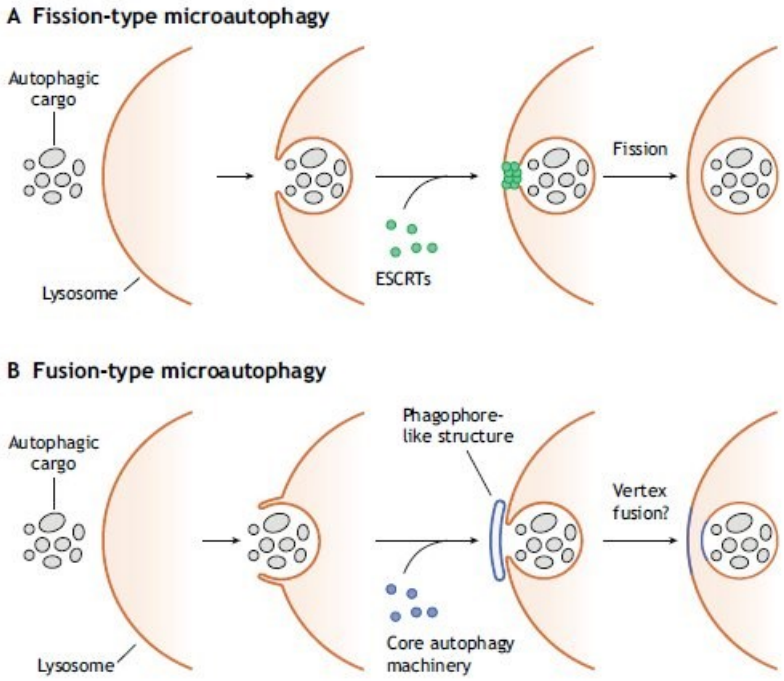


Figure 14. Schematic representation of the two type of microautophagy (Schuck, 2020).

Fission-type requires proteins called endosomal sorting complex required for transport (ESCRTs) for the membrane scission to happen, while SNAP receptors (SNAREs) are not required for generating a new membrane structure. This type of microautophagy is peculiar of general microautophagy and for degradation of the ER, both in mammals and yeast and the specificity of the substrate is granted by the direct binding between the ubiquitinated cargoes and the ESCRT proteins or by facilitators such LC3 and Hsc70.

Fusion-type microautophagy, on the contrary, requires SNAREs and the core autophagy machinery such as Atg30 and Atg39 which ensure selectivity to the cargoes [Schuck, 2020].

1.6.3 Macroautophagy

Macroautophagy, widely simply referred as autophagy, is a process involved in many aspects of the cell life, from the protection from stresses like starvation to the normal cellular development, senescence, immunity and preservation against microbial incursions. Moreover, macroautophagy is altered in many human pathologies, such as cancer, neurodegeneration, skeletal muscle disorders, heart, live and gastrointestinal diseases.

It goes without saying that autophagy is one of the most tightly regulated processes in which any fluctuation in excess or defect, could deleteriously affect the cell survival.

While initially believed to be a non-specific process, triggered by the nutrient deprivation, it is nowadays well known that macroautophagy could also be highly specialized in order to maintain a proper cellular homeostasis.

The activation of the macroautophagy pathway involves with the generation of a phagophore or isolation membrane which is the precursor of the double-membrane vesicles called autophagosomes. This last can engulf large cytoplasm sections or organelles and will eventually fuse with a lysosome to form an autolysosome. In this single membrane organelle, the cargoes are degraded by the lysosomal lytic enzymes and the product of degradation will then be recycled into the cytoplasm through permeases [Klionsky et al, 2014; Galluzzi et al, 2017].

Autophagosome formation requires different steps and a complex molecular machinery that was identified thanks to the extensive work of Ohsumi, which showed the astonishing conservation and homology between the yeast and the mammalian macroautophagy machinery.

The first step in the autophagosome formation is the *induction*, regulated in yeast by the Atg1 complex constituted by Atg1, Atg11, Atg13, Atg17, Atg29 and Atg 31 proteins. In mammals, the induction requires the assembly of the homolog ULK1 complex containing, beside ULK1 or ULK2, the mammalian homologs ATG13, ATG101, RB1 inducible coiled-coil 1 (RB1CC1 also known as FIP200).

At the putative phagophore assembly site (PAS), is then summoned for *nucleation*, the ATG14-containing class III phosphatidylinositol 3-kinase (PtdIns3K) complex, consisting of PIK3C3/VPS34, PIK3R4/VPS15, BECN1, the nuclear receptor binding factor 2 NRBF2 and the membrane curvature sensor ATG14.

The *elongation* of the phagophore is regulated by two systems equally present in yeast and mammals, the ATG12– ATG5-ATG16L1 complex which dissociates from the autophagosome after completion and the Atg8/LC3 complex that in mammals is composed of subfamilies with different isoforms (LC3 family including LC3A, LC3B, LC3B2 and LC3C; GABARAP family (including GABARAP, GABARAPL1 and GABARAPL2). LC3 is lipidated with the addition of phosphatidylethanolamine (PE) by ATG4, giving rise to a cytoplasmic LC3-I and its PE-conjugated counterpart LC3-II which is integrated into the autophagosome's membranes.

The elongated double membrane will finally enclose autophagy substrates and close to form a *mature* autophagosome which will eventually fuse with a lysosome becoming an autolysosome. The trafficking and fusion of the autophagosome toward the lysosome is dependent on

microtubules and it possibly involves components of the endocytic SNARE machinery [Chen et al, 2011; Parzych et al, 2014; Galluzzi et al, 2017] (Fig.15).

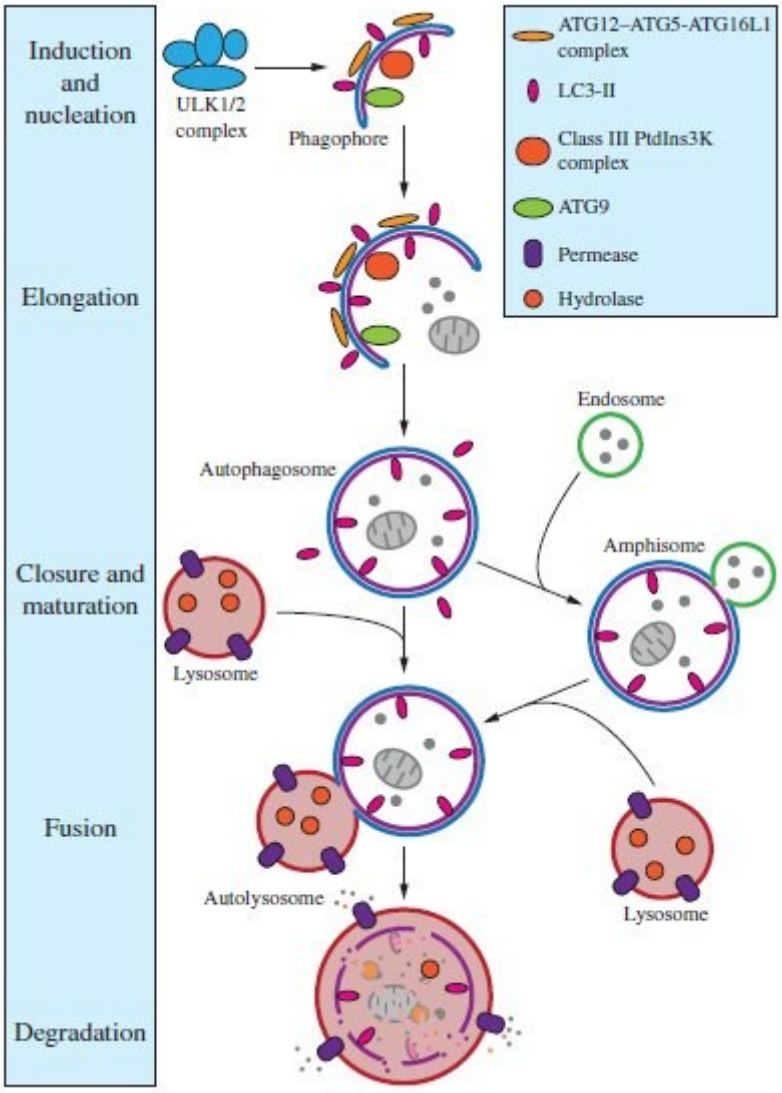


Figure 15. Representation of the fundamental steps and their regulator in the formation of mature autolysosomes (Parzych et al, 2014).

This process of autolysosome formation is highly regulated and one of the most effective regulator is MTOR complex 1 (mTORC1) that, by phosphorylating ATG13 and ULK1 inhibits the autophagosome formation. The inhibiting action of mTORC1 is also carried out by preventing TFEB (a transcriptional regulator of lysosome biogenesis) from translocating outside the nucleus.

mTORC1 activity is itself is inhibited the AMP-activated protein kinase (AMPK), which senses the reduction in ATP levels [Jung et al, 2009; Settembre et al, 2012; Galluzzi et al, 2017].

As said, beside the non-specific engulfment of portions of the cytoplasm, the macroautophagy is also greatly specialized being able to degrade diverse organelles and compartments so that specific macroautophagy processes can be identified.

These are defined as:

- mitophagy, the degradation of damaged as well as perfectly functioning but superfluous mitochondria.
- nucleophagy, the selective degradation of nucleus portions, also contributing to the maintenance of the genomic integrity
- pexophagy, the targeting of peroxisomes upon ubiquitination of several proteins of the peroxisomes that are then recognized by the autophagy receptors p62/SQSTM1 and NBR1.
- ribophagy, the degradation of ribosomes, so far only described in yeast and some plants, yet to be reported in mammals.
- lipophagy, the degradation of Lipid Droplets (LDs) which in mammals is transcriptionally regulated.
- bacterial and viral xenophagy, the removal of bacteria that have escaped phagosomes upon phagocytosis or the cytoplasmic formed virions.
- proteophagy, the disposal of inactive proteasomes.
- lysophagy, the elimination of damaged lysosomes.
- aggrephagy, the selective removal and degradation of protein aggregates.

1.7 Aggrephagy

Proteins that undergo misfolding, form oligomeric complexes interacting with other partially or unfolded proteins, thus producing insoluble protein aggregates that become gradually larger and accumulate within the cell cytoplasm. These aggregates can arrange into histologically variable structures such as intracellular inclusions, bodies, tangles or threads and can be amorphous or structured such the amyloids [Fink, 1998; Kopito, 2000].

These insoluble aggregates, even though ubiquitinated, are not fitting substrates for the narrow barrel-shaped proteasome unless they are partially unfolded. Therefore, another degradation mechanism must be activated, that is a specialized form of macroautophagy named aggrephagy by Selgen and its group in 2007 [Øverbye et al, 2007] (Fig.16).

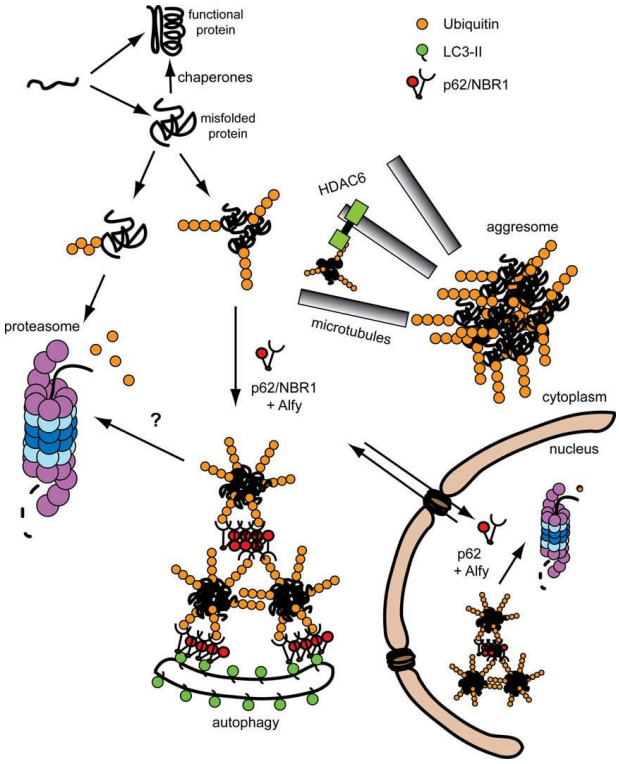


Figure 16. Schematic representation of the mechanisms related to the degradation of misfolded ubiquitinated proteins forming aggregates (Knaevelsrud & Simonsen, 2010).

The initial formation of aggregates is possibly happening co-traslationally, with an almost immediate aggregation in small particles of these misfolded proteins. A mechanism of retrograde transport based on dynein is then moving these forming particles toward the microtubule organization center (MTOC). Here, larger particles are produced by fusing with each other, thus defining the assembly of a structure called aggresome. These structures are generally enclosed in a mesh of the intermediate filament vimentin, even though this doesn't appear strictly necessary for the aggresome deposition, since aggregates formed by α-synuclein are not surrounded by this cage [Garcia-Mata et al, 2002].

More misfolded proteins are also recruited to the forming aggresomes by the histone deacetylase 6 (HDAC6).

The aggresomes have been found to be associated with chaperones such as members of the Hsc70 family, ubiquitinating enzymes and components of the proteasome [Garcia-Mata et al, 2002; Lamark et al, 2012].

Seen the stability of such aggresomes and the overwhelming burden for the proteasome, the macroautophagy machinery is then recruited for degradation, assisted by specialized proteins such as the autophagy receptor p62/SQSTM1, the protein next-to-BRCA1 gene 1 (NBR1), optineurin

(OPTN), toll-interacting protein (TOLLIP) and the WD repeat and FYVE domain containing 3 protein (WDFY3 also known as Alfy).

1.7.1 Histone deacetylase 6 (HDAC6)

Histone deacetylase 6 (HDAC6), is a tubulin deacetylase localizing to the cytoplasm, that possess a ubiquitin-binding domain (BUZ domain) and a cytoplasmic dynein binding domain located between two deacetylase domains. Therefore, HDAC6 can bind simultaneously to the ubiquitinated aggregated proteins (preferentially to the K63 of ubiquitin) and to dynein, thus serving as a linker for the retrograde transport of protein aggregates [Kawaguchi et al, 2003; Iwata et al, 2005] (Fig.17).

Moreover, cells lacking HDAC6 show the inability to bind to polyubiquitinated aggregates and to transport them via the dynein, causing an impairment of the aggregates clearance from the cytoplasm resulting in a stress-induced apoptosis [Kawaguchi et al, 2003].

It was shown in 2010 by Lee and coworkers that HDAC6 is necessary for the fusion of the autophagosomes with the lysosomes. This fusion is promoted by the recruitment of cortactin (a component of the F-actin polymerization machinery), which allows the production of a meshwork of F-actin at the aggresomes site [Lee et al, 2010].

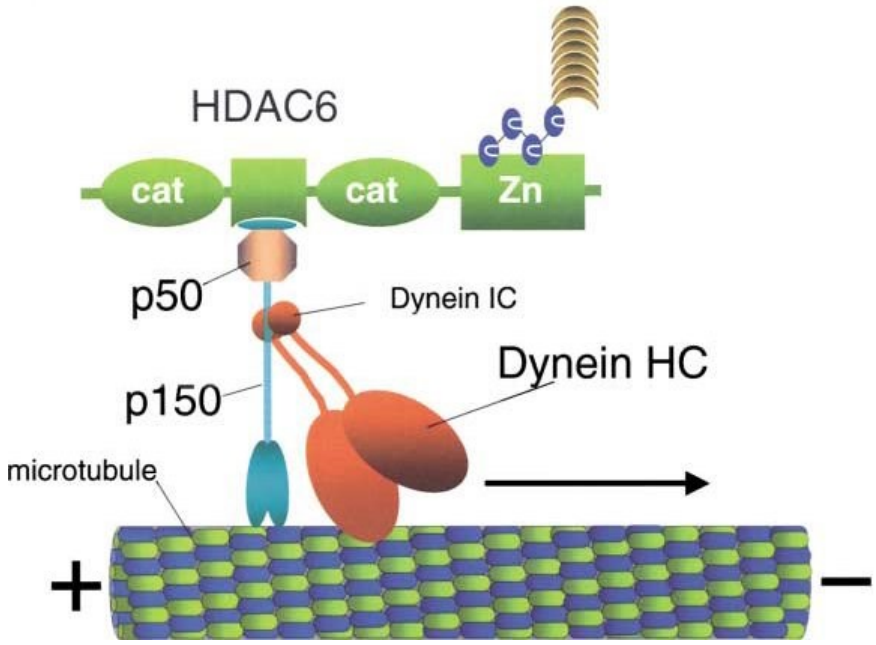


Figure 17. Illustration of the association between HDAC6 and dynactin/dynein, allowing the retrograde movement of protein aggregates toward the MTOC (p150 and p50 dynactin's subunits). (Modified from Kopito, 2003).

1.7.2 p62/SQSTM1

Sequestosome 1 (SQSTM1), also known as p62 is a protein involved in many aspects of autophagy. Moreover, it is directly involved in the recognition and targeting of the aggregates for degradation via macroautophagy.

It has been shown the presence of p62/SQSTM1 in neurofibrillary tangles of tauopathies, synucleinopathies and Alzheimer's disease patients' brains [Kuusisto et al, 2001; Kuusisto et al, 2002].

p62/SQSTM1 possess a ZZ Zinc finger domain, an oligomerization domain (PB1), a LIR domain, and a C-terminal UBA domain.

The *LIR domain* of p62/SQSTM1, allows its interaction with the Atg8 mammalian homologues, the LC3/GABARAP family members. Fundamental consensus motif for this interaction it's a 22 amino acid sequence (Arg^{321}-Ser^{342}) that allows the direct contact with the hydrophobic pocket of the ubiquitin domain [Pankiv et al, 2007]. It was subsequently shown by Komatsu's group that an 11 amino acid sequence is sufficient for this bond to happen (Ser^{334}-Ser^{344}) [2008_Ichimura]. Moreover, this interaction requires both the N and C termini of the LC3 protein. It is also via this LIR motif that p62 is itself degraded and recycled through macroautophagy [Bjørkøy et al, 2005; Pankiv et al, 2007; Ichimura et al, 2008].

Another pivotal region of p62/SQSTM1 is the C-terminal *Ubiquitin associated (UBA) domain*. Through this domain, p62 is able to bind non-covalently ubiquitin and polyubiquitin, thus allowing the connection with proteins that have been targeted for degradation by ubiquitination [Vadlamudi et al, 1996; Donaldson et al, 2003; Cavey et al, 2005]. Remarkably, phosphorylation at Ser^{403} by Casein Kinase 2 (CK2), strengthen p62/SQSTM1 affinity for polyubiquitin chains [Matsumoto et al, 2011] (Fig.18). Vice versa, when p62 inactivation is required, the UBA domain is homodimerized, thus preventing p62/SQSTM1 ability to interact with ubiquitin [Long et al, 2010; Isogai et al, 2011]. Mutations in the UBA domain have been associated with Paget's disease of bone (PDB) [Cavey et al, 2005].

Moreover, via its *PB1 domain*, p62/SQSTM1 interacts with NBR1 and is able to form head-to-tail helical filaments that will be the gathering sites for ubiquitinated proteins and for the assembly of protein aggregates [Bjørkøy et al, 2005; Ichimura et al, 2008]. This domain is also crucial for degradation of p62/SQSTM1 itself through autophagy [Ichimura et al, 2008].

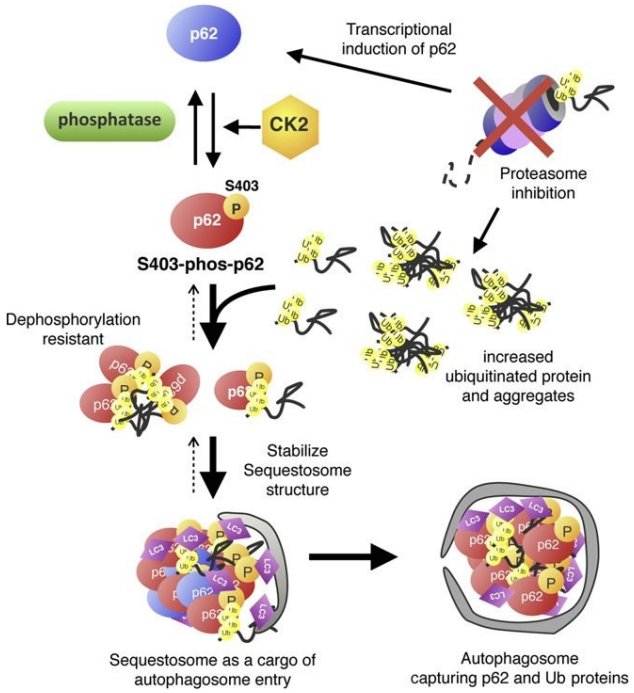


Figure 18. p62/SQSTM1 is usually in equilibrium between a phosphorylated and dephosphorylated state. When ubiquitinated proteins accumulate in the cytoplasm the phosphorylated p62/SQSTM1 binds these aggregates. The phosphorylated p62/SQSTM1 conjugated with ubiquitinated proteins is resistant to dephosphorylation and can thus contribute to the formation of sequestosomes which can then be sequestered into isolation membranes summoned by the LC3- p62/SQSTM1 interaction (Matsumoto et al, 2011).

1.7.3 Next to BRCA1 gene 1 (NBR1)

In 2009 Kirkin and colleagues identified a new autophagy receptor found in complex with p62/SQSTM1 [Kirkin et al, 2009a]. This new protein next-to-BRCA1 gene 1 (NBR1), is larger than p62/SQSTM1 and has poor similarity at the amino acid level with p62/SQSTM1, but it conserves the structural domains that are present in its counterpart. It is in fact characterized by a LIR, UBA and PB1 domains (Fig.19).

NBR1 directly interacts and cooperates with p62/SQSTM1 as cargo adapter for protein aggregates that need to be targeted to the autophagosomes for degradation.

NBR1, similarly to p62/SQSTM1, has been found in synucleinopathies localizing in the Lewy bodies (LBs) and glial cytoplasmic inclusions (GCIs) [Odagiri et al, 2012]. Furthermore, NBR1 along with p62/SQSTM1, ubiquitin and phosphorylated tau, has been found to accumulate in the muscle fibers of patients with sporadic inclusion body myositis (sIBM) [D'Agostino et al, 2011].

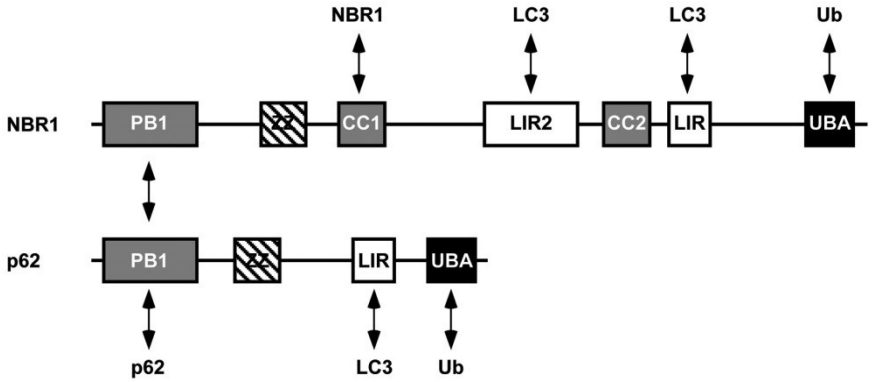


Figure 19. Schematic representation showing the comparison between the structural domains of NBR1 and p62/SQSTM1. (Kirkin et al, 2009b).

1.7.4 Autophagy-linked FYVE (Alfy)/WDFY3

Alfy also known as WD repeat and FYVE domain-containing protein 3 (WDFY3), is a large protein of 400 kDa ubiquitously expressed and containing at its C-terminal region a PH- BEACH domain, five WD40 repeats and a FYVE domain, a zinc finger that mediates the interaction with phosphatidylinositol 3-phosphate (PtdIns3P) (Fig.20) [Simoensen et al, 2004] .

Furthermore, Simonsen's group, identified the presence of a LIR sequence in Alfy's WD40 region (aa 2981-3526), that is responsible for the interaction with GABARAP subfamily [Lystad et al, 2014].

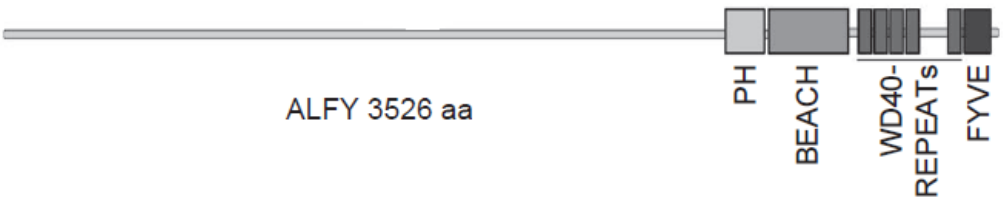


Figure 20. Schematic representation of the large Alfy protein, showing the domain's distribution at the C-terminus of the protein (modified from Lystad et al, 2014).

It was shown that Alfy and p62/SQSTM1 interact via the PH-BEACH domain and form a complex, co-localizing within the cytoplasmic p62/SQSTM1 bodies [Clausen et al, 2010], pinpointing Alfy as necessary for the elimination of protein aggregates through autophagy (but not for starvation induced autophagy), most likely serving as a scaffold protein (Fig.21) [Filimonenko et al, 2010].

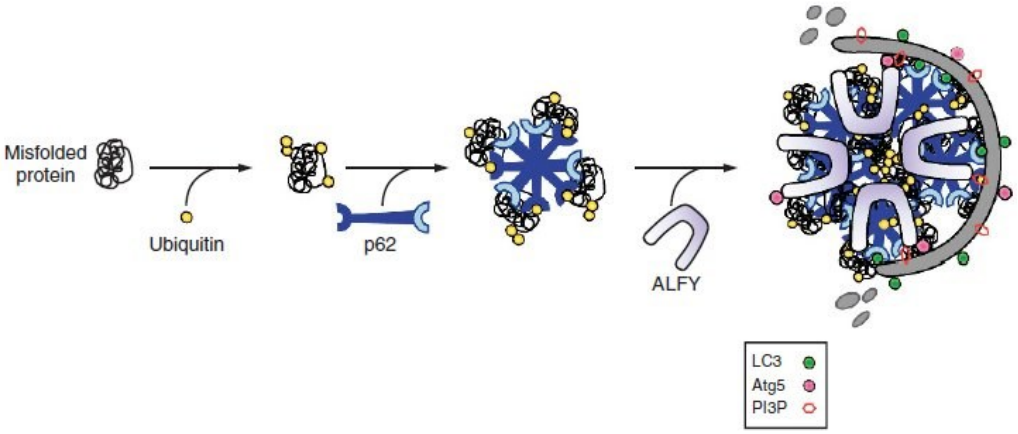


Figure 21. Possible mode of action of Alfy in the aggrephagy process. The ubiquitinated proteins are linked to p62/SQSTM1 which forms aggregates by homopolymerizing via its PB1 domain. Alfy is then beckoned from the nucleus to the p62/SQSTM1/ubiquitin-positive cytoplasmic structures, facilitating the formation of larger aggregates that can be degraded by autophagy (modified from Isakson et al, 2013).

2. Aim of the study

During my working activity at the Muscle Cell Laboratory at the IRCCS Fondazione Istituto Neurologico Carlo Besta in Milan, a large family affected by a dominant distal vacuolar myopathy with late onset and slow progression came to our attention.

The clinical symptoms overlapped with numerous other distal myopathies but the haplotype shared among all the affected members of the family was excluding any other, so far reported, distal myopathies.

Focus of my research for the past few years has been the identification of the genetic cause underlying this disorder. The genetic analysis proved to be challenging, with many NGS approaches that were attempted, going from exome, to genome as well as RNA sequencing. None of these approaches allowed us to identify a possible causative mutation.

The mystery was only solved when a proteomic approach was first applied, analyzing the content of the numerous vacuoles present in the muscle fibers of the patients.

We then identified the causative mutation as being a large coding repeat expansion in the exon 3 of *PLIN4* gene, encoding for the amphipatic helix of the perilipin 4 protein. Another challenge has now begun consisting in the elucidation of the molecular mechanisms leading to this vacuolar muscular disorder, possibly through activation and imbalance of the aggrephagy pathway.

Aim of my PhD project has been the identification and characterization of this unique mutation causing a muscle disorder.

3. Material and Methods

3.1 Patients

Patients had signed the informed consent to use biological samples for diagnosis and research. The study was conducted in accordance with protocols approved by the institutional review boards of the C. Besta Neurological Institute of Milan.

Clinical features showed progression of the symptoms over time with initial weakness of scapular muscles, extending to the distal lower limb muscles. The levels of creatine kinase were normal or slightly increased. All affected individuals, including the asymptomatic one had an electromyography (EMG) characterized by myopathic potentials in proximal and distal limb muscles and frequent repetitive high frequency discharges.

The asymptomatic patient (V:13) (see Fig.22) examined at the age of 32 years revealed, on the Medical Research Council (MRC) scale, hand extensor (4/5), neck flexor (4/5) and hallux extensor (4/5) muscle weakness. Her electromyography complied with the pattern present in the symptomatic patients showing positive repetitive complex discharges at needle insertion and myopathic pattern, particularly in tibialis anterior and biceps brachialis muscles. The biopsy of quadriceps muscle was characterized by some centralized nuclei and vacuoles in three myofibers.

3.2 Morphological studies

Quadriceps muscles from patients IV:3, IV:10, IV:21, IV:23, IV:24, V:13, IV:12 and IV:17 were snap-frozen in isopentane/liquid nitrogen, and then processed for morphological studies. Transverse cryostat sections were routinely stained for: haematoxylin-eosin, Gomori trichrome, periodic acid-Schiff, Oil Red O, myofibrillar ATPase, acid phosphatase, NADH dehydrogenase, cytochrome C oxidase, succinic dehydrogenase. Primary antibodies for immunofluorescence recognized: poly-ubiquitinylated proteins (monoclonal PW8810, clone FK2 from Biomol, Enzo Life Sciences, Inc. Farmingdale, USA; 1:100), p62/SQSTM1 (guinea pig polyclonal GP62-c, Progene Biotechnik, Heidelberg, Germany; 1:100), p62/SQSTM1 (rabbit polyclonal P0067, Merck, Darmstadt, Germany 1:100) Perilipin 4 (guinea pig polyclonal GP34, Progene Biotechnik, Heidelberg, Germany; 1:100), NBR1 (mouse monoclonal 4BR, Santa Cruz Biotechnology Inc, Santa Cruz CA, USA; 1:100), WDFY3 (mouse monoclonal B-4, Santa Cruz Biotechnology Inc, Santa Cruz CA, USA; 1:100). Primary antibodies were applied to the cryosections for 2hr, followed by 1hr secondary incubation with an anti-guinea pig, or anti-mouse IgG, Alexa 555 or Alexa 488 conjugated (Invitrogen Life

Technologies, Carlsbad, CA, USA; 1:1500). Muscle sections were examined under either a Zeiss Axioplan fluorescence microscope (Carl Zeiss AG, Oberkochen, Germany) or a Leica confocal microscope equipped with hybrid and argon lasers (Leica Microsystems, Wetzlar, Germany). Small fragments of muscle tissues, fixed in 4% glutaraldehyde and embedded in Epon 812 resin (Electron Microscopy Sciences, Hatfield PA, USA), were stained with UranyLess (Electron Microscopy Sciences, Hatfield PA, USA) and examined under a FEI Technai electron microscope.

3.3 Linkage analysis

A combination of Illumina chip arrays was used to generate data for whole genome linkage analysis. In detail, SNPs from Illumina Omni1 chip formerly used for CNV analysis on two patients (IV:10 and IV:17) were combined with SNPs from Illumina Omni2.5 array run for patients III:18, IV:3, IV:23 and V:13. Linkage dysequilibrium-based SNP pruning on communal SNPs was then performed using PLINK [Purcell et al, 2007] and MERLIN software (version 1.1.2) [Abecasis et al, 2002] was used for parametric linkage analysis with parameters considering a rare, dominant and fully penetrant disorder (0.0001, 1.0, 1.0). To further define the exact boundaries of the haplotype shared among the affected members of the family, we combined our previous microsatellite analysis results with the SNPs array data.

3.4 Genome sequencing

Genome sequencing was performed at The Centre for Applied Genomics in Toronto, following Illumina TruSeq Nano DNA library preparation protocol. The sequencing of the equimolar pooled libraries was performed on an Illumina HiSeq X platform following Illumina's protocol, generating paired-end reads of 150-bases in length.

3.5 RNA-seq

RNA-seq on two samples from affected patients was done at The Centre for Applied Genomics in Toronto using the NEBNext Ultra Directional RNA library prep kit for Illumina with poly(A) enrichment. Libraries pooled in equimolar quantities were pair-end sequenced on an Illumina HiSeq 2500 platform following Illumina's protocol generating paired-end reads of 126-bases in length.

3.6 Laser microdissection, protein digestion and mass spectrometry

Frozen sections from two affected samples (IV:12 and IV:21) were collected on a specific support and by laser microdissection using LMD 6500 (Leica Microsystems), different tissue's areas were collected. The areas were selected as following: rimmed vacuoles (defined as VAC), an equivalent portion of the sarcolemma of the vacuolated fibers (defined as PCI, patient control I), and an equivalent portion of the sarcolemma of healthy-looking fibers (defined as PCII, patient control II). The digested samples were then submitted to nano high-performance liquid chromatography-mass spectrometry (nanoHPLC-ESI-MS/MS) [Maerkens et al, 2016].

3.7 Relative protein quantification

The data obtained by ESI-MS/MS analysis mass spectrometry were matched against the entire Uniprot/Swissprot (release 2014/10, 546,790 entries) database. For each protein, the ratios between the averaged proportions in rimmed vacuoles and control samples were calculated and statistical significance was evaluated with a two-tailed unpaired t-test (equal variances assumed).

3.8 Western Blot analysis

Fragments of the muscle tissues were homogenized with a buffer containing 20 mM Tris (pH 7.8), 137 mM NaCl, 2.7 mM KCl, 1 mM MgCl2, 1% Triton X-100, 10% (w/v) glycerol, 1 mM EDTA, 1 mM dithiothreitol supplemented with protease inhibitor cocktail (P8340, Merck, Darmstadt, Germany) and incubated 1h at 4°C. After a 10 min centrifugation at 13,000 rpm at 4°C, the supernatant was collected and protein content was assayed with the DC protein kit (Bio-Rad, Hercules, California, USA). Total proteins were run on a 7.5% or 4-15% acrylamide gels, then transferred into a nitrocellulose membrane for immunological detection using as primary antibodies: perlipin-4 (guinea pig polyclonal AP33136SU-N, OriGene, Rockville, USA) and SDH70 (anti-complex II 70 kDa Fp subunit monoclonal antibody, cod 459200, Invitrogen, Thermo Fisher Scientific, Carlsbad, California, USA) as loading control.

3.9 Exon 3 length evaluation by PCR amplification

Segregation was performed in all the available genomic DNA by amplification with Phire Hot Start II DNA polymerase (Thermo Fisher Scientific, Carlsbad, CA, USA) with primers 3F-ACCACTCGGTCTGCACTTA and 3R-CCTCCAACTCATTCTGCAGC. The PCR products were run on a 1%

agarose gel identifying the presence of a single (wt) or two (wt and mut) bands with a UV transilluminator equipped with a Kodak EDAS 290 camera (Eastman Kodak Company, Rochester NY, USA).

3.10 Oxford Nanopore Technologies long read sequencing

Nanopore sequencing was performed on 500ng of linearized plasmids (pCMV6-AN-GFP) containing either the wild type or the mutant *PLIN4* allele of patient cDNA retrotranscribed from total muscle RNA extract. Both samples were prepared for sequencing with the 1D Ligation Sequencing Kit (SQK-LSK108, Oxford Nanopore Technologies) using the manufacturer's recommendations. Base calling and demultiplexing were performed with Guppy version 3.2.4 (Oxford Nanopore Technologies). Demultiplexed reads were then subjected to de novo assembly using Canu version 1.8 [Koren et al, 2017]. The contig assembled from the most reads was then aligned to the human reference genome DNA sequence (GRCh38/hg38).

3.11 Plasmids generation

For the generation of the full-length protein plasmid construct, muscle RNA from an affected patient was extracted using TRIzol reagent (Thermo Fisher Scientific, Carlsbad, CA, USA) and retrotranscribed using the High Capacity RNA-to-cDNA kit (Thermo Fisher Scientific, Carlsbad, CA, USA). The product was amplified with the Phire Hot Start II DNA polymerase (Thermo Fisher Scientific, Carlsbad, CA, USA) in order to add the necessary restriction sites for cloning into the pCMV6-AN-GFP plasmid (OriGene, Rockville. MD, USA), using the following primers: FW-TAAGAGGCGATCGCCAtgcagaccctgggcagcttctttgggtc and RW-TGCGCGACGCGTctactgcccgccagcgggcaaggcgaagg. The PCR product was then run on a 1% agarose gel, both bands were excised and column purified (Macherey-Nagel, Düren, Germany).

The two bands as well as the commercial plasmid, were digested with the SgfI and MlUI restriction enzymes, further agarose gel and column purification was performed and the final products quality was verified on a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Carlsbad, CA, USA). The linearized plasmid and the PCR products were then ligated using the T4 ligase from Promega (Promega, Madison, WI, USA) according to the manufacturer's indications. One Shot TOP10 Chemically Competent E. coli cells (Thermo Fisher Scientific, Carlsbad, CA, USA) were transformed with the ligation products and grown ON at 37°C on amplicillin plates. The following

morning, grown colonies were picked, further expanded on liquid LB and minipreps were prepared to test for positive colonies (Macherey-Nagel, Düren, Germany). Positive colonies showing the presence of the wt or the mut alleles were kept and glycerol stocks were stored.

3.12 Cell transfection

The day before transfection, 2 x 10^4 HeLa cells were plated in each 24 well/plate with cover glasses of the appropriate diameter, in a medium containing 10% FBS and no antibiotics. The following day, four mixes were obtained combining 1 µg of the specific plasmid with 2 µl of Lipofectamine 2000 (Thermo Fisher Scientific, Carlsbad, CA, USA) in the appropriate amount of OPTIMEM medium. Each mixture was supplemented to the established wells. Oleic acid (Merck, Darmstadt, Germany) was added at the final concentration of 250 µM to induce the formation of LDs. Transiently transfected cells were kept for 24 or 48 h, then fixed with 4% paraformaldehyde and mounted on glass slides for confocal microscope analysis.

Stable cell lines were obtained by adding the Neomycin (G418 Merck, Darmstadt, Germany) antibiotics at the final concentration of 400 µM, thus selecting for those cells that had incorporated the construct.

4. Results

4.1 Family presentation

4.1.1 Pedigree

For several years, we have been investigating a large Italian family affected by a dominant distal myopathy with late onset and slow progression of the symptoms (Fig.22).

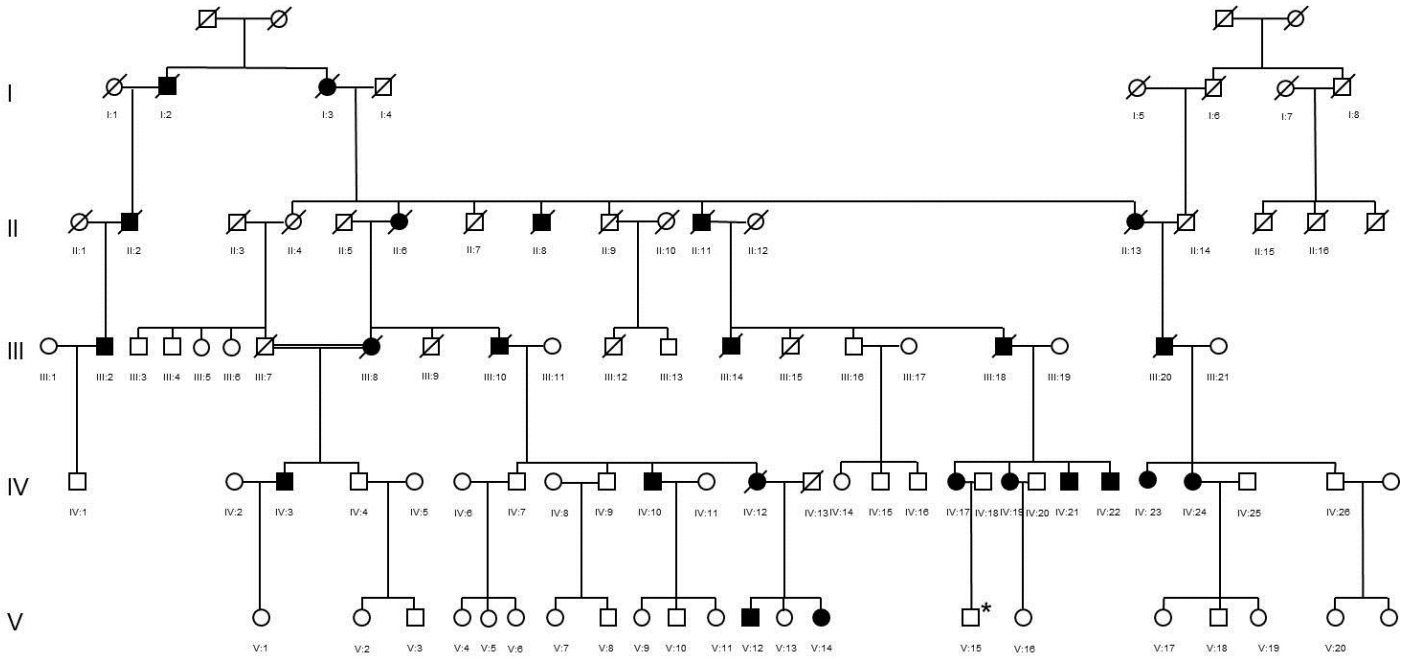


Figure 22. Family tree showing the segregation of the disease over five generations. Circles represent females, squares males. Black circles and squares are affected patients. Barred circles and squares represent deceased individuals. The star indicates an asymptomatic patient.

4.1.2 Clinical traits

The first signs of the disorder are generally manifesting around the third or fourth decade of life (40 ± 5 years) and the severity of the symptoms increases with time.

In the very early phases of the disease the patients might only show a very subtle impairment of the strength of the finger extensors and neck flexor with inability to walk on heels, but no other disturbances reported.

Nevertheless, the end of the spectrum manifests with critical weakness of the bilateral foot dorsiflexor and neck flexors, as well as with involvement of the shoulders and fingers' muscles, without fasciculation. The final outcome for these patients is the use of a wheelchair with just a hinted contraction of abductors and flexors of the upper limbs but not of the lower limbs.

No patients at any time ever reported dysphagia or dystonia.

The cardiac muscle and respiratory functions also seem to be spared, always exhibiting preserved performances.

In all affected patients, pre-symptomatic as well as severely affected, the electromyography reveals positive repetitive complex discharges at needle insertion and small MUPs from all muscles and particularly from tibialis anterior and biceps brachialis.

The values of creatine kinase (CK) proved to be doubled compared to the control, in the most severe cases, yet within the normal range in the least severe patients (reference values 30-150 U/L).

The muscle imaging shows atrophy and substitution by fat tissue in the most severe patients with no visible alterations in the least affected ones (Fig.23).

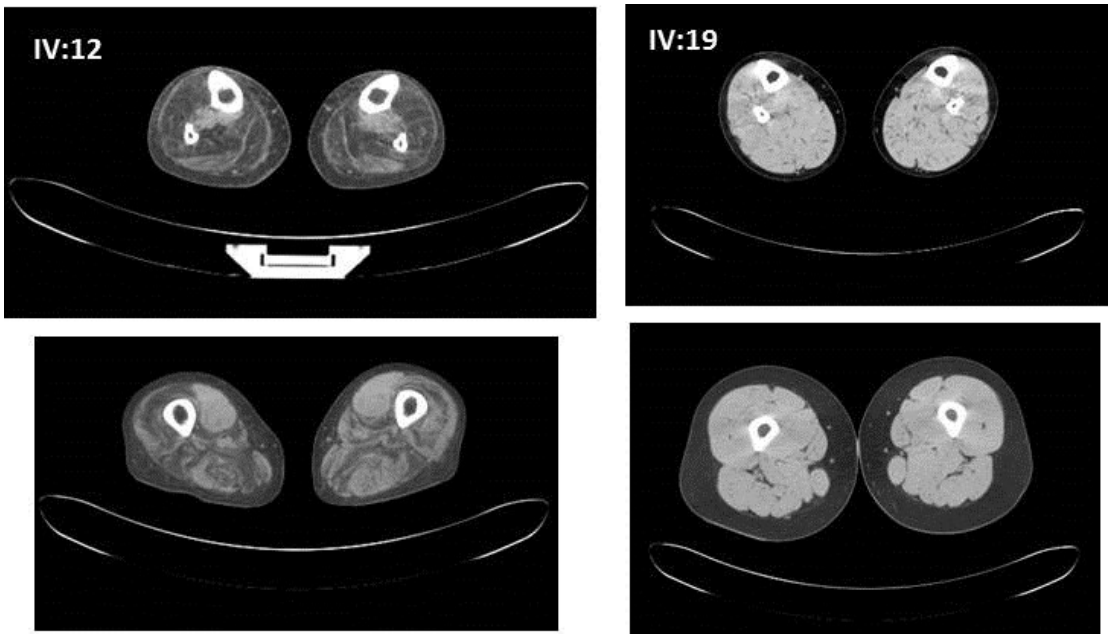


Figure 23. Computerized axial tomography (CAT) scans of the lower limbs of the more severely affected patient IV:12 and the very mild patient IV:19. Noticeable, the significant substitution of the muscle tissue in patient IV:12, while muscle in patient IV:19 is preserved.

4.1.3 Histological hallmarks

Histology was performed on needle biopsies of quadriceps muscles of affected patients. Myopathic features are hypertrophic, normotrophic and hypotrophic fibers, fiber splitting, angulated fibers, degenerated fibers and centralized nuclei. In the least severe patients these features are less pronounced but still noticeable (Fig.24). Rimmed vacuoles of variable sizes, within or at the surface of the fibers are also visible. The vacuoles contain partially degraded organelles, and amorphous material, as visible by electron microscopy (Fig.25), but are always negative to Thioflavin-S and Congo red staining, indicating absence of β-amyloid deposits. The signals for ubiquitin and proteasome's subunits 19S and 20S are highly increased in the subsarcolemmal region of some fibers as well as within the vacuoles. The lipid content, as revealed by the Oil Red O (ORO) staining, is comparable to a normal control muscle (Fig.26).

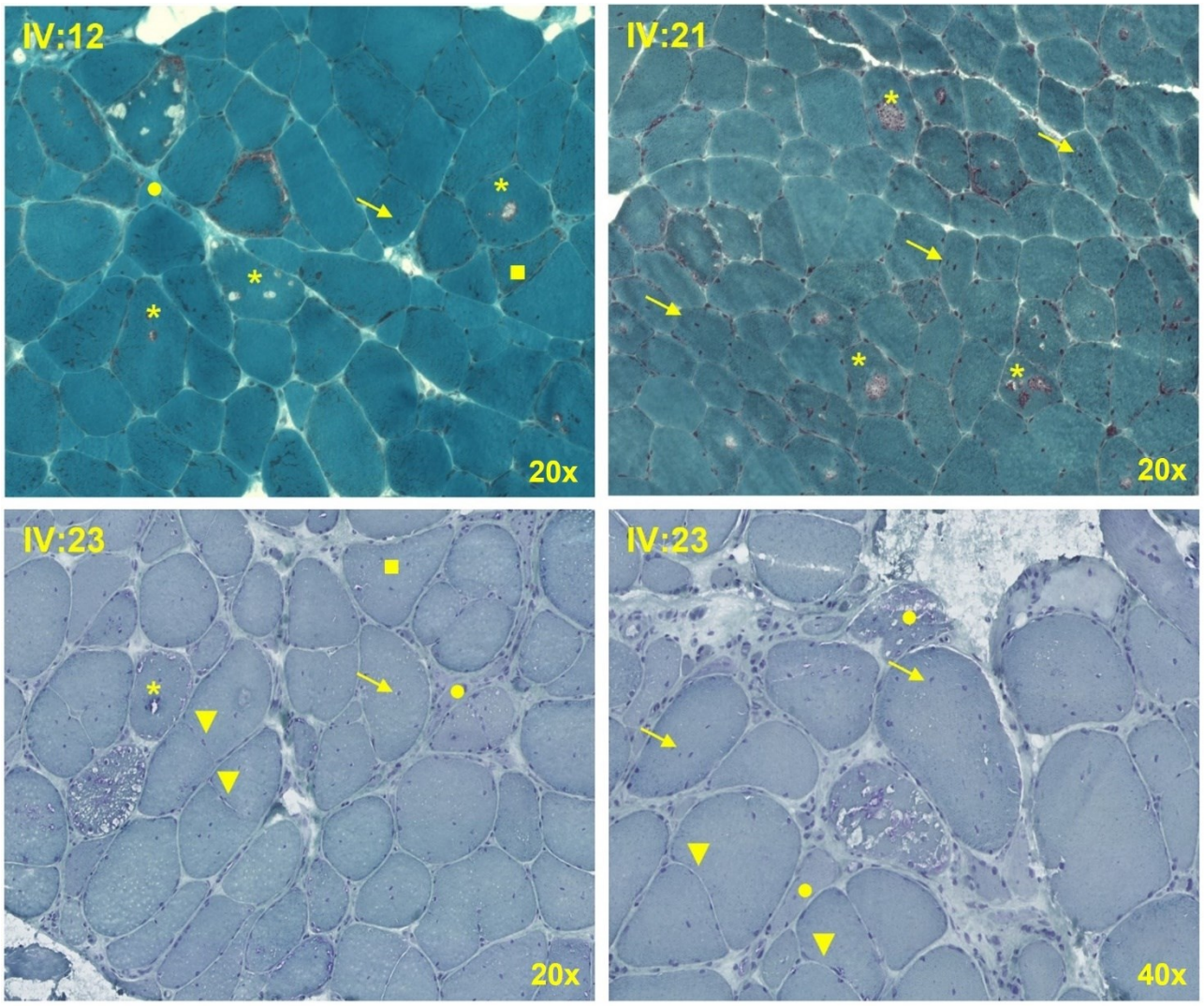


Figure 24. Representative histology images of affected quadriceps muscles (patients IV:12, IV:21 and IV:23). Variable fiber sizes, with angulated fibers (squares), degenerating fibers (dots), fiber splitting (arrow heads), central nuclei (arrow) and numerous vacuoles (stars) are easily recognizable.

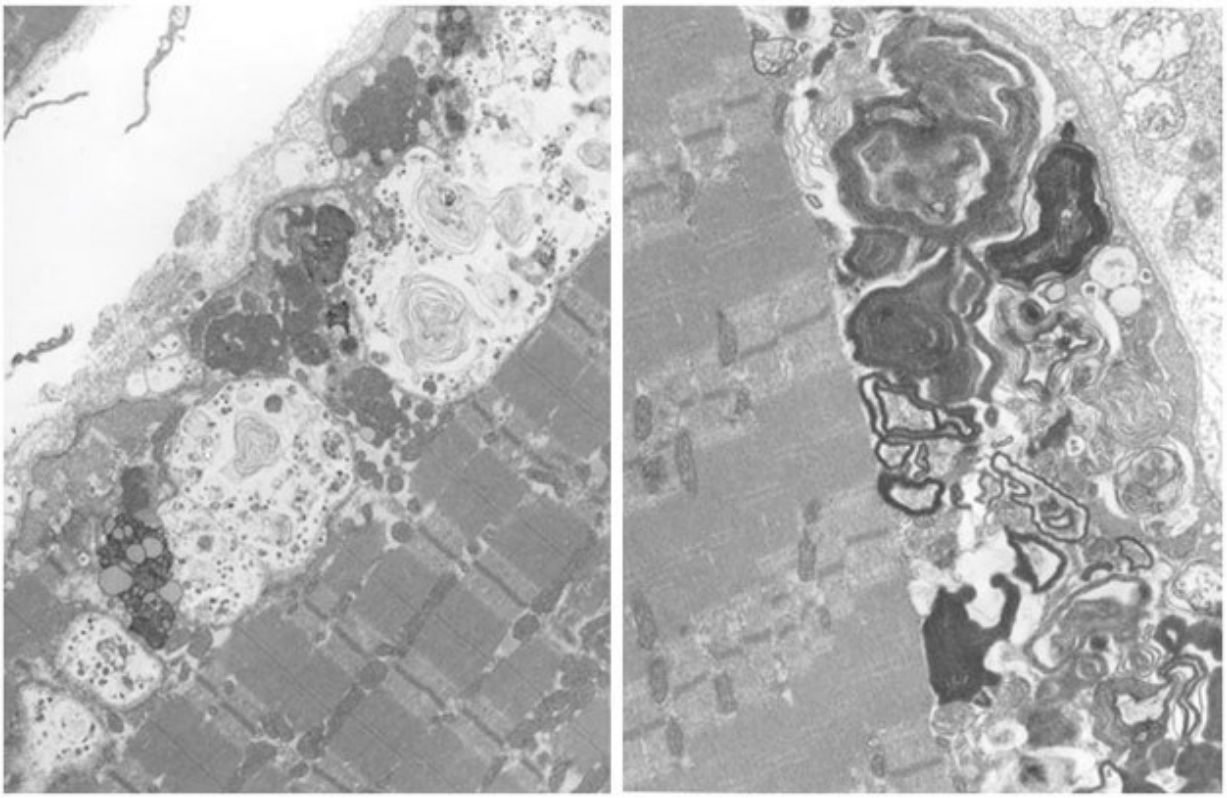


Figure 25. Electron microscopy of vacuoles in patient IV:12. Vacuoles are visible at the surface of the fibers, containing swirls, myelin figures, degraded organelles and amorphous material.

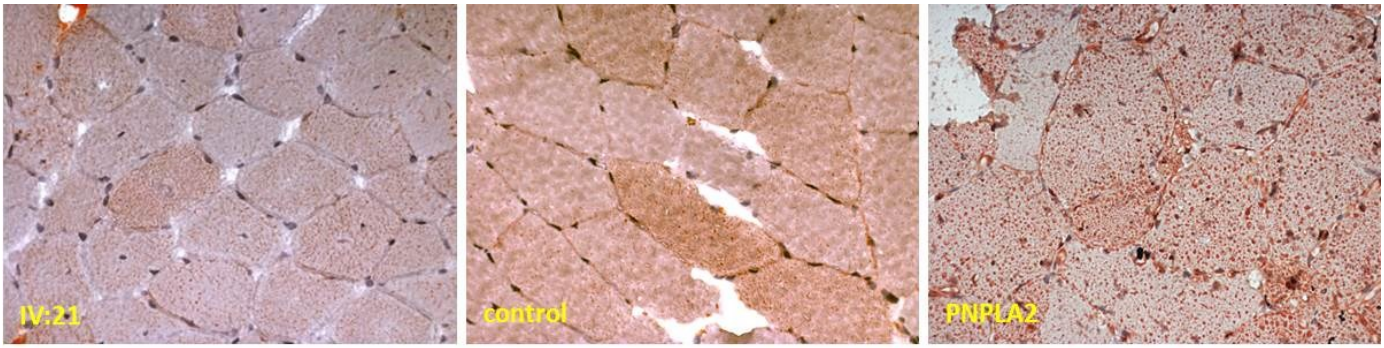


Figure 26. Oil red O staining on the muscle tissue. The distribution of the lipids within the muscle fibers (red dots within the fibers) is comparable between patient (IV:21) and a control muscle. In a PNPLA2 patient, with a mutation that impairs the function of the adipose triglyceride lipase (ATGL), the number of accumulating lipid droplets is markedly increased.

4.2 Previous results: linkage refinement, WGS and RNA sequencing, p62/SQSTM1 and FK2 immunohistochemical characterization

In a young girl (V:14), daughter of an affected patient (IV:12), we identified a chromosome recombination within the haplotype region of chromosome 19, by microsatellite analysis. She underwent neurological, electromyography and muscle biopsy examinations and was thus diagnosed as affected.

Therefore we carried out a new linkage analysis on DNA samples from six individuals including this new patient (IV:10, IV:17, III:18, IV:3, IV:23 and V:14), combining the information from the microsatellites and a SNP array from Illumina. We could thus refine the disease haplotype to a 5,12 Mb region, containing 164 genes.

Seen the considerable number of genes present in this region, Sanger sequencing was not a feasible option, thus we selected two affected and two unaffected siblings for whole genome sequencing, that was performed at The Centre for Applied Genomics (TCAG) in Toronto. We carried out the analysis of the WGS data and after filtering for frequency, correct segregation and dominant inheritance pattern, we were left with over 1000 intronic or intergenic variants of unknown significance.

In order to clarify some of these variants, we performed RNA-seq from the skeletal muscle of two affected patients. The RNA-seq data revealed no aberrant expression of genes in the linked region, nor obvious functional correlations between these genes and genes with altered expression elsewhere in the genome.

While reviewing the NGS data, we performed immunohistochemical analysis on all the muscle biopsies available, using two antibodies, the first recognizing p62/SQSTM1 and the second identifying the ubiquitinated proteins (clone FK2). We choose to study these markers since the aberrant number of vacuoles within the muscle fibers is correlated with altered autophagic activities. We observed an increased signal for both antibodies at the subsarcolemmal region of the fibers and within the vacuoles, unlike in any other vacuolar myopathy that we are aware of. The signals were strikingly co-localizing and there was an undeniable correlation between the severity of the phenotype and the number of fibers and vacuoles that were positive for these antibodies, ranging from few in the young lady V:14 to all fibers in the most severely affected patients (Fig.27).

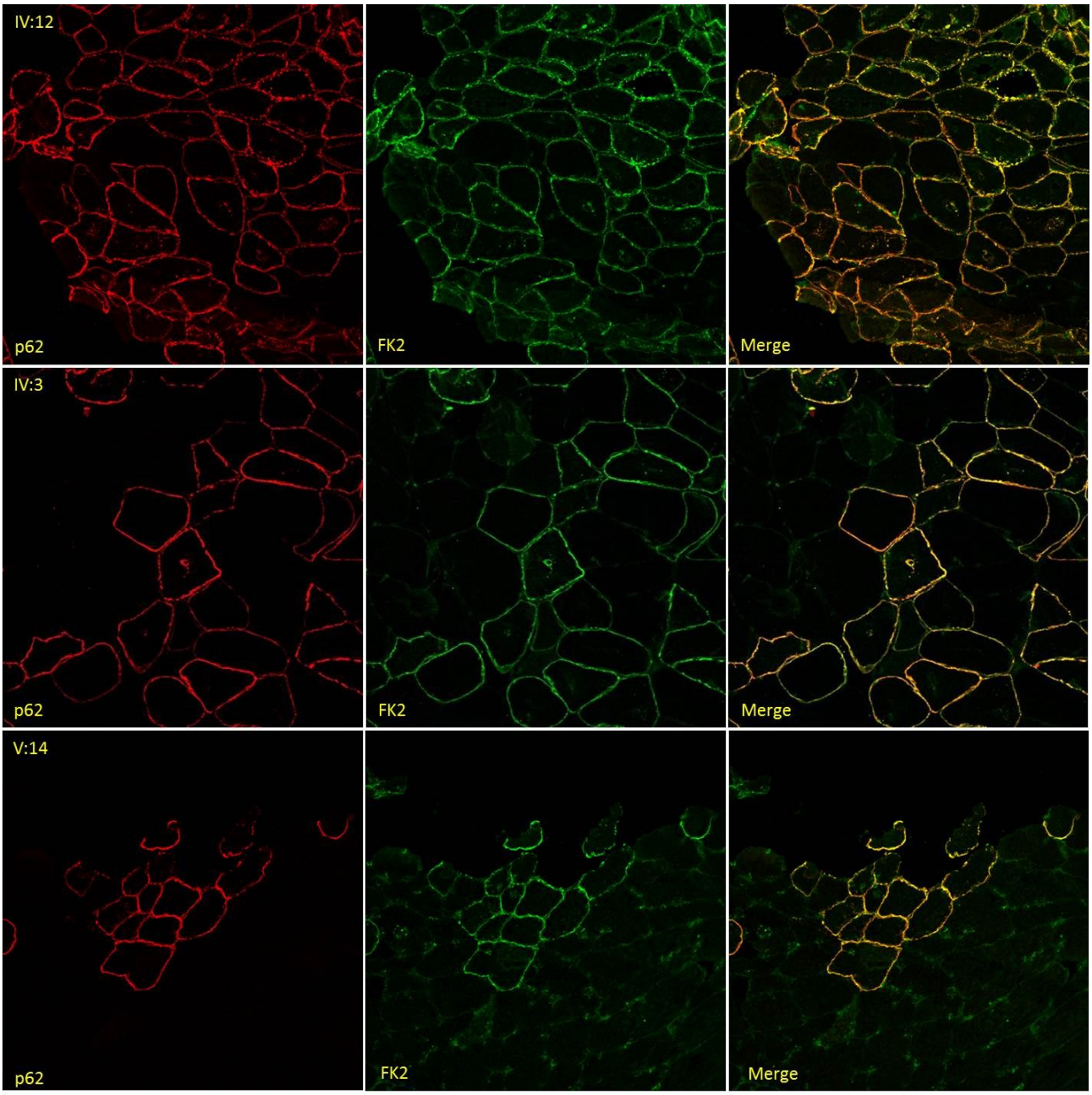


Figure 27. Immunofluorescence analysis using antibodies recognizing p62/SQSTM1 and the ubiquitinated proteins (FK2). Both signals are localizing in the subsarcolemmal region of some fibers and within the vacuoles. Remarkable is the different level of positivity for both antibodies that directly correlates with the clinical severity.

These findings prompted us to hypothesize that the protein that was ubiquitinated and that was recognized by p62/SQSTM1, could have been our disease protein, seen the perfect correlation between its labeling and the severity of the symptoms.

Therefore we sent one of the affected muscle biopsy for laser microdissection of the vacuoles and subsequent mass spectrometry to identify the proteins accumulating within the vacuoles.

4.3 Laser microdissection and mass spectrometry

The analysis was carried out in collaboration with the group of Dr. Kley (Department of Neurology, Neuromuscular Centre Ruhrgebiet, Ruhr-University Bochum, Bochum, Germany). Slices of the affected muscle biopsies were prepared on special supports and vacuoles (label VAC) were cut out and collected for mass spectrometry analysis. The content of these vacuoles was compared with equivalent portions of the sarcoplasm of the same vacuolated fiber (PCI, patient control I), with a portion of a healthy looking fiber from the same biopsy (PCII), or of a healthy control muscle fiber (HC) (Fig.28).

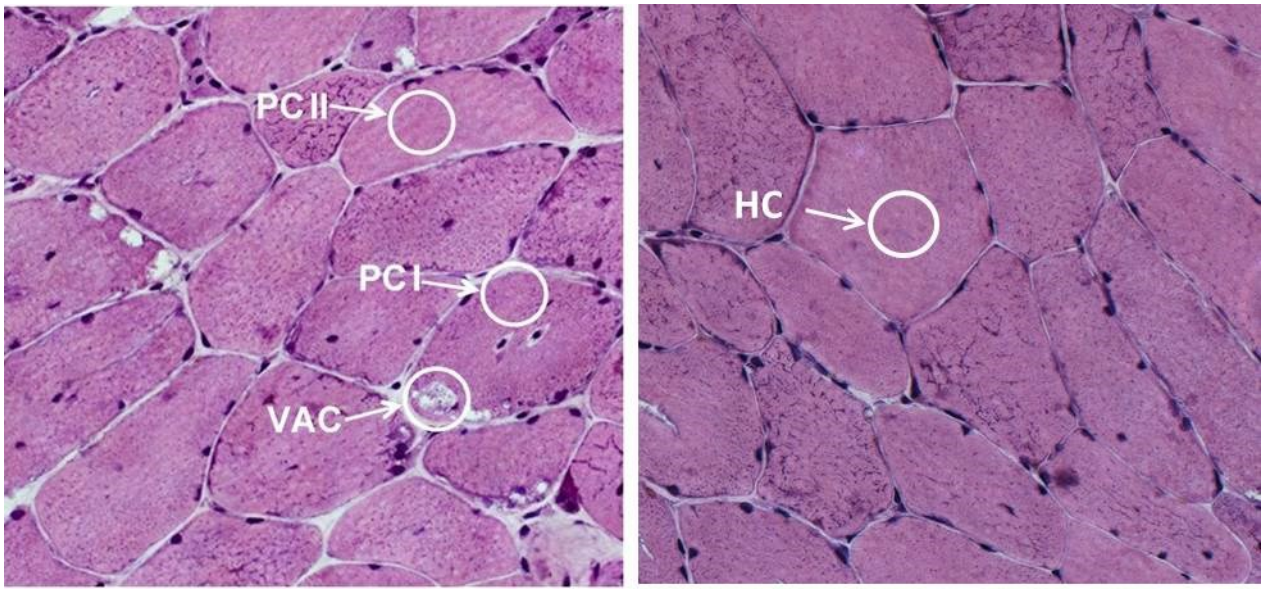


Figure 28. Representation of the different areas that were compared in the mass spectrometry analysis.

The results of the mass spectrometry analysis revealed a list of over 700 proteins that were up or down regulated. We then focused our attention on the over represented proteins in the patient's sample and among these, we looked for proteins encoded by genes in the region of the shared haplotype on chromosome 19. The only protein fulfilling both requirements was perilipin 4 that was 20 fold more represented in the vacuoles of the patient compared to same areas of control samples and whose gene is indeed located in the shared region (Fig.29).

Within the vacuoles, other proteins were found to be upregulated. However, these proteins were muscle proteins such as dystrophin or desmin that were expected to be found within the vacuoles. Moreover, none of these proteins were as much represented as perilipin 4.

Accession	Locus	Protein names	Ratio VAC / HC	TTEST	VAC ID1	VAC ID2	HC ID1	HC II ID2	mean VAC	mean HC II
Q96Q06	chr19:4502192-4517716	Perilipin-4 (Adipocyte protein S3-12)	19,57080292	0,135548422	6,812652068	3,012048193	0	0,502008032	4,91235013	0,251004016

Accession	Locus	Protein names	Ratio VAC / PC II	TTEST	VAC ID1	VAC ID2	PC II ID1	PC II ID2	mean VAC	mean PC II
Q96Q06	chr19:4502192-4517716	Perilipin-4 (Adipocyte protein S3-12)	19,57080292	0,135548422	6,812652068	3,012048193	0,502008032	0	4,91235013	0,251004016

Accession	Locus	Protein names	Ratio VAC / PC I	TTEST	VAC ID1	VAC ID2	PC I ID1	PC I ID2	mean VAC	mean PC I
Q96Q06	chr19:4502192-4517716	Perilipin-4 (Adipocyte protein S3-12)	0,850904475	0,741803026	6,812652068	3,012048193	7,02811245	4,518072289	4,91235013	5,773092369

Accession	Locus	Protein names	Ratio PC I / PC II	TTEST	PC I ID1	PC I ID2	PC II ID1	PC II ID2	mean PC I	mean PC II
Q96Q06	chr19:4502192-4517716	Perilipin-4 (Adipocyte protein S3-12)	23	0,049744732	7,02811245	4,518072289	0,502008032	0	5,77309237	0,251004016

Figure 29. Proteomic data showing the results for perilipin 4. Ratios of perilipin 4 in vacuoles and the different controls areas are shown.

4.4 Perilipin 4 in patients

4.4.1 Genetic validation

Since *PLIN4* maps to our linked region, we revisited the genomic and transcriptomic patient data specifically for *PLIN4* gene, but again could found no significant point mutations. However, we noticed an unusually high sequencing coverage in the exon 3 of the gene, a peak that was consistently increased in the RNA-seq data and in the WGS data and that was absent in the healthy siblings as well as in a cohort of other patients processed at the TCAG facility (Fig.30).

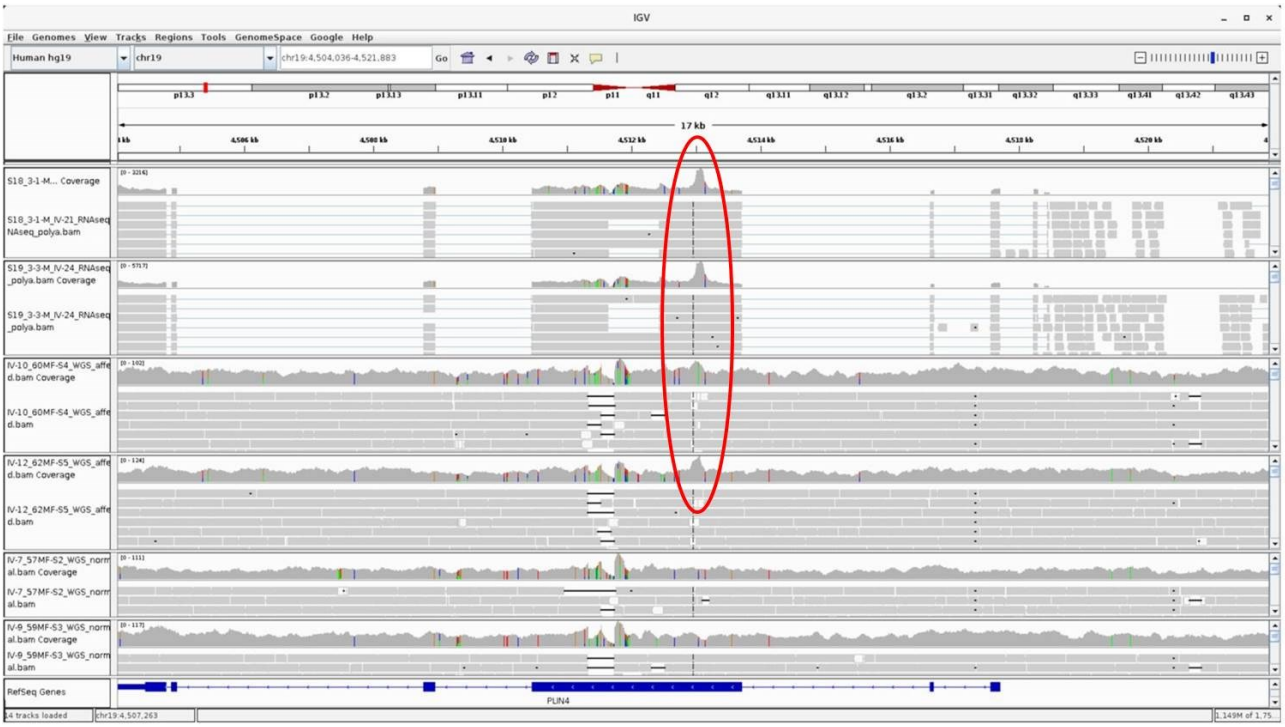


Figure 30. IGV (Integrative Genomics Viewer, Robinson et al, 2019) visualization of Next Generation Sequencing results, both RNA (rows 1, 2) and genome (rows 3-6) data, showing presence in patients of a high peak of sequencing coverage (rows 1-4) in PLIN4 exon 3 not present in the healthy siblings (row 5, 6).

This peak spanned a region of roughly 120 bp in exon 3 that we decided to amplify by PCR from the patient's DNA. Seen the considerable repetitiveness nature of the entire exon 3, in order to design primers specific for unique genomic sequences, we had to amplify the entire exon.

Surprisingly, the amplification of this exon in the patient's genomic DNA and muscle RNA, revealed the presence of the wild-type band as well as a second band that appear to be ~1000 bp larger compared to the wild-type. This higher band was present in all the affected patients but was never found in the healthy relatives nor in 60 ethnically matched control individuals (Fig.31).

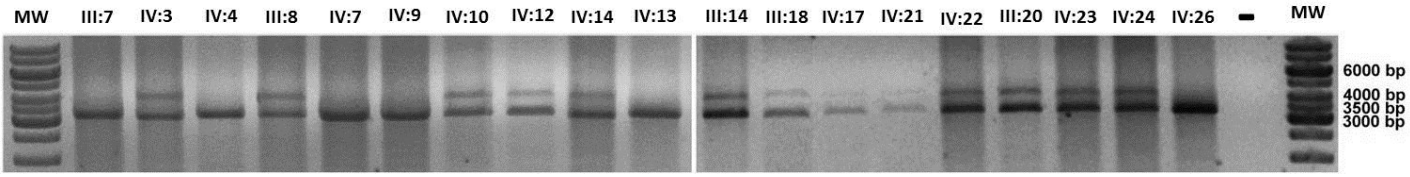


Figure 31. The PCR amplification of the entire exon 3 shows that, in addition to the wild-type band, a second band approximately 1000 bp larger is present in affected but not unaffected members of the family.

The amphipathic domain of perilipin 4 is structurally organized in imperfect repeats and is completely encoded by exon 3, reflecting the protein organization in a nucleotide sequence characterized by a high degree of repetitiveness. This feature makes the Sanger sequencing not feasible, but also poses a computational challenge for aligning short sequencing reads produced by the mainstream NGS techniques.

Therefore we decided to analyze our DNA samples with a long range sequencing approach based on Oxford Nanopore Technology.

The first sequencing attempt on the genomic DNA from one affected patient, revealed to be ineffective, because the coverage in the haplotype region was not adequate and many artifacts were introduced during the sequencing.

Therefore, because of the complexity of the *PLIN4* gene sequence and the high error rate that is intrinsic in the Oxford Nanopore Technology, we choose to separate the two bands amplified by PCR, corresponding to the entire exon 3, and analyzed them separately by long range sequencing to enrich our samples and reduce unspecific signals.

The Oxford Nanopore Technology sequencing data revealed that the mutant allele presents an expansion of the three identical amino acid 33-mer that are present in this region. Therefore, rather than having 3 x 99 nt array as in the wild-type band, the mutant allele presents 12 x 99 nt asset resulting in 891 extra nucleotides (9 x 99) and 297 extra amino acids (9 x 33). This expansion

results in a perilipin 4 isoform that possess an amphipatic region of 40 x 33-mer instead of the canonical 31 x 33-mer (Fig.32).

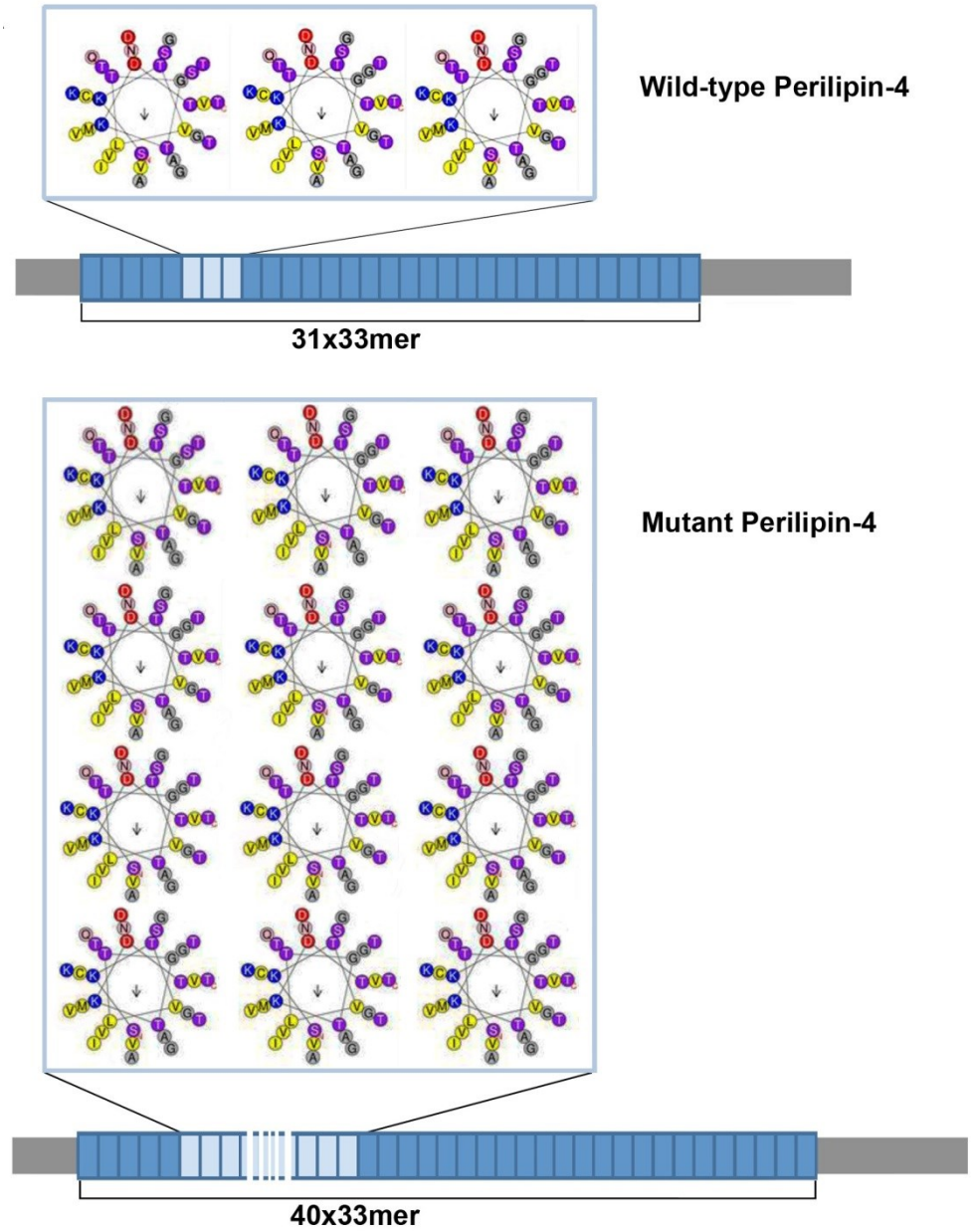


Figure 32. Graphical representation of the protein expansion caused by the extra 891 nucleotides in exon 3 of *PLIN4* gene. The expanded identical 33-mer corresponds to amino acids 246-278, 279-311 and 312-344 in the wild-type sequence.

4.4.2 Immunohistochemical validation

Since perilipin 4 is reported to be present in skeletal muscle and to be localized at the subsarcolemmal region of the fibers [Pourteymour et al, 2015], we investigated the expression of perilipin 4 in our patients' muscle biopsies.

Using an antibody recognizing the N-terminal region of the protein, we found a major increase of perilipin 4 positivity in the subsarcolemmal regions of the fibers and within the vacuoles, when compared to control biopsies.

Moreover, perilipin 4 signal almost exactly reproduced the staining observed with the FK2 and p62/SQSTM1 antibodies (Fig.33) while neatly recapitulating the correlation between number of fibers with increased signal and phenotypic severity (Fig.34).

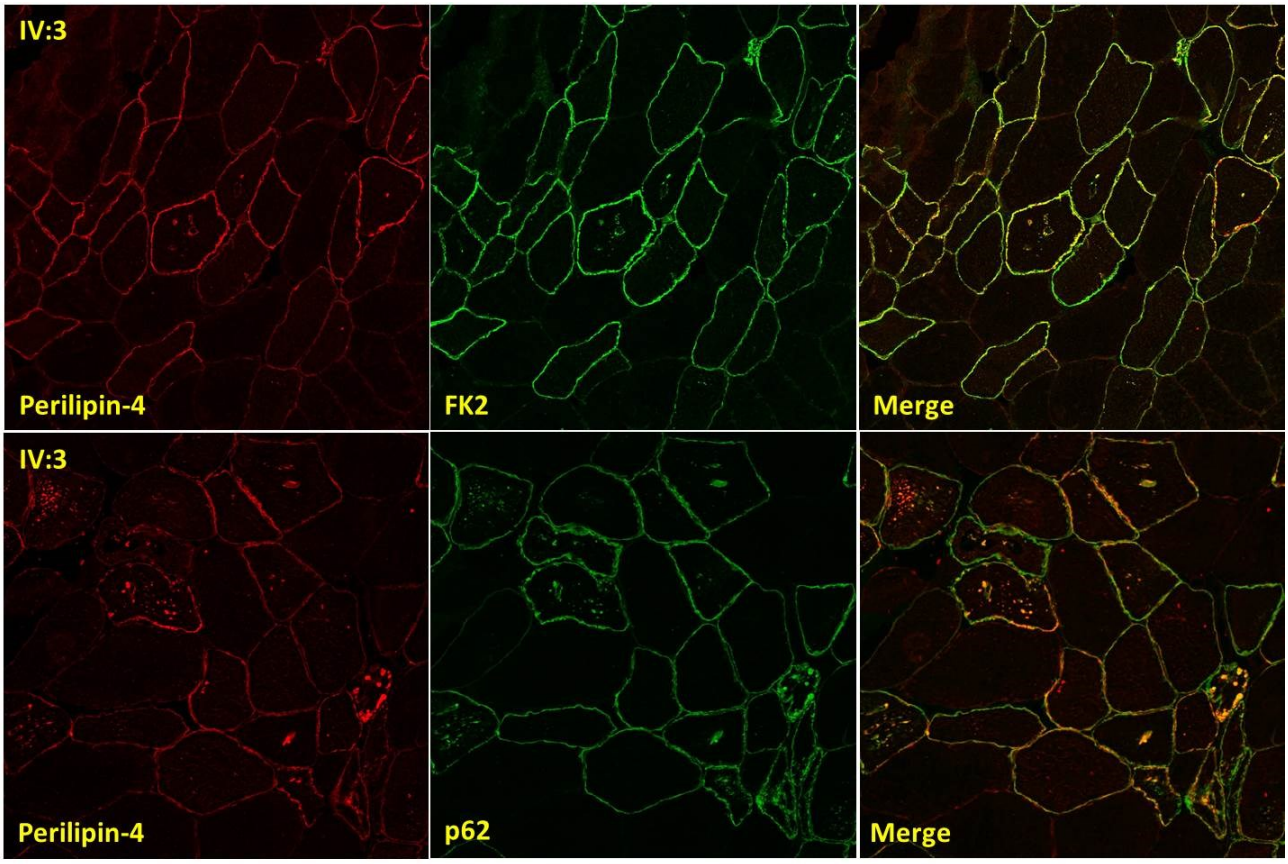


Figure 33. Double staining of patients' muscle biopsies using the antibodies recognizing perilipin 4 and the previously used FK2 (upper panel) and p62/SQSTM1 (lower panel). There is an almost perfect overlap between these antibodies, both at the subsarcolemmal region and within the vacuoles.

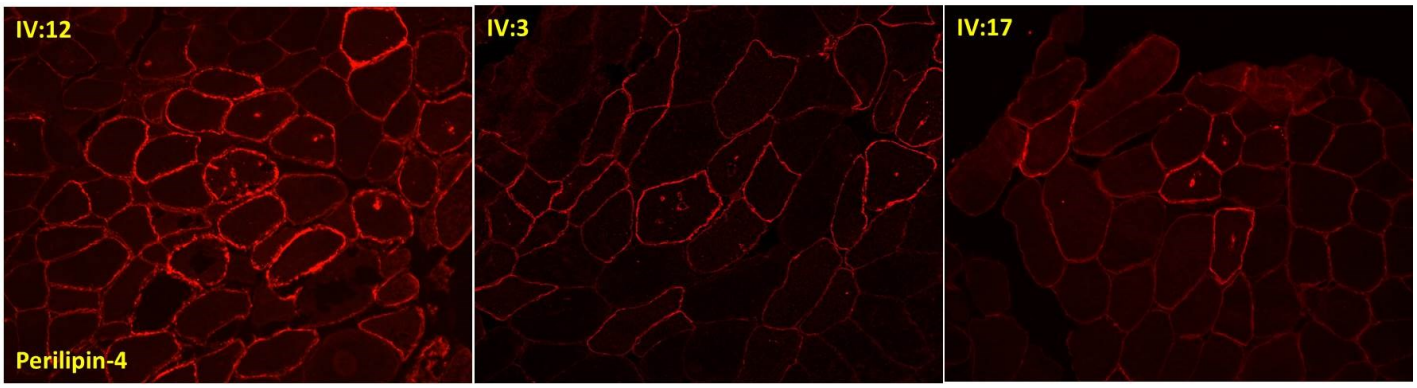


Figure 34. Immunostaining on patients' muscle biopsies, showing the localization of perilipin 4 in the subsarcolemmal region of the fibers and within the vacuoles. Moreover, it is evident the decreased positivity to the perilipin 4 antibody, going from the most severe patient (IV:12) to the least severe one (IV:17).

4.4.3 Western blot validation

Since the expansion in the *PLIN4* gene is within a coding region, we were expecting to observe the presence of a higher isoform, along with the wild-type size one, in the protein extract obtained from the muscle biopsy.

Indeed, by western blot analysis, we identified the presence of a second band of a size consistent with the genetic expansion in patients, which was absent in control samples. Particularly, the expected size for the wild-type isoform protein is about 130 kDa, while the mutant isoform is anticipated to be approximately 30 kDa higher than the previous one. These values are consistent with the western blot findings (Fig.35).

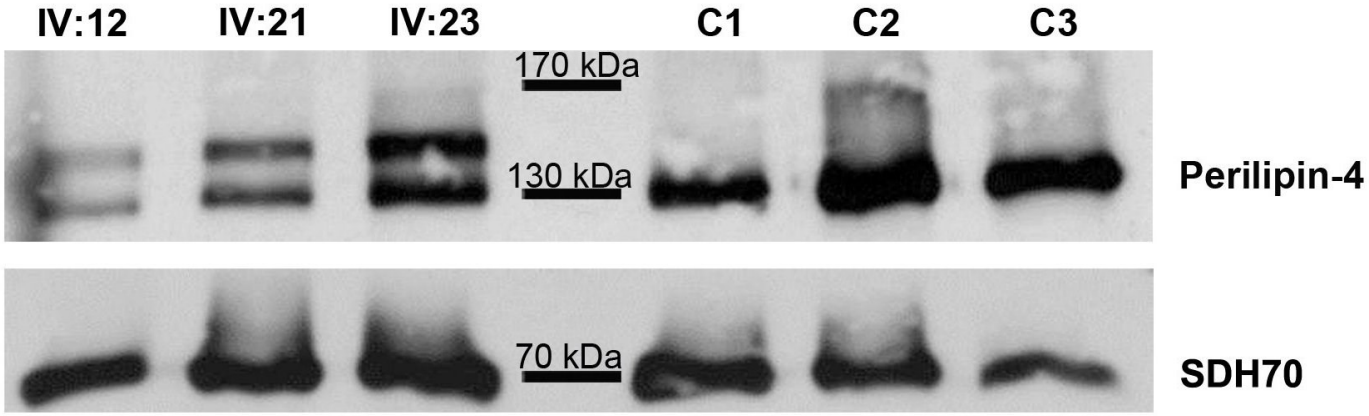


Figure 34. Western blot analysis of the muscle biopsies protein extracts. In all affected patients, but not in the healthy controls, it is visible a second band higher than the wild-type, with an apparent size that is consistent with the identified genetic insertion. The control used is the 70 kDa subunit of the succinate dehydrogenase (SDH70).

4.5 Evaluation of the aggrephagy pathway by immunohistochemical analysis

Having verified the existence of this alternative longer isoform of perilipin 4, we were left with a puzzling question: which is the mechanism that links a protein involved in the coating of lipid droplets to a distal vacuolar myopathy?

Following the indications suggested by the p62/SQSTM1 and FK2 analysis, we reasoned that the aggrephagy pathway might have been involved, thus justifying the presence of the numerous vacuoles that are filling the fibers of the most severely affected patients.

Circling back to our patient muscle biopsies, we then examined some of the components of the aggrephagy pathway by immunofluorescence analysis.

We observed that, beyond FK2 and p62/SQSTM1, other components of the pathway were upregulated, namely NBR1 (Neighbor of BRCA1) and WDFY3 (also known as Alfy).

Particularly, NBR1 which is known to directly interact with p62/SQSTM1 via the PB1 domain, is co-localizing with perilipin 4 (Fig.35), FK2 and p62/SQSTM1, while WDFY3, a large protein believed to form a scaffold for the localization of aggresomes, is increased in the subsarcolemma near perilipin 4 positivity without co-localization (Fig.36). Moreover, both positivity directly correlates with the severity of the phenotypes.

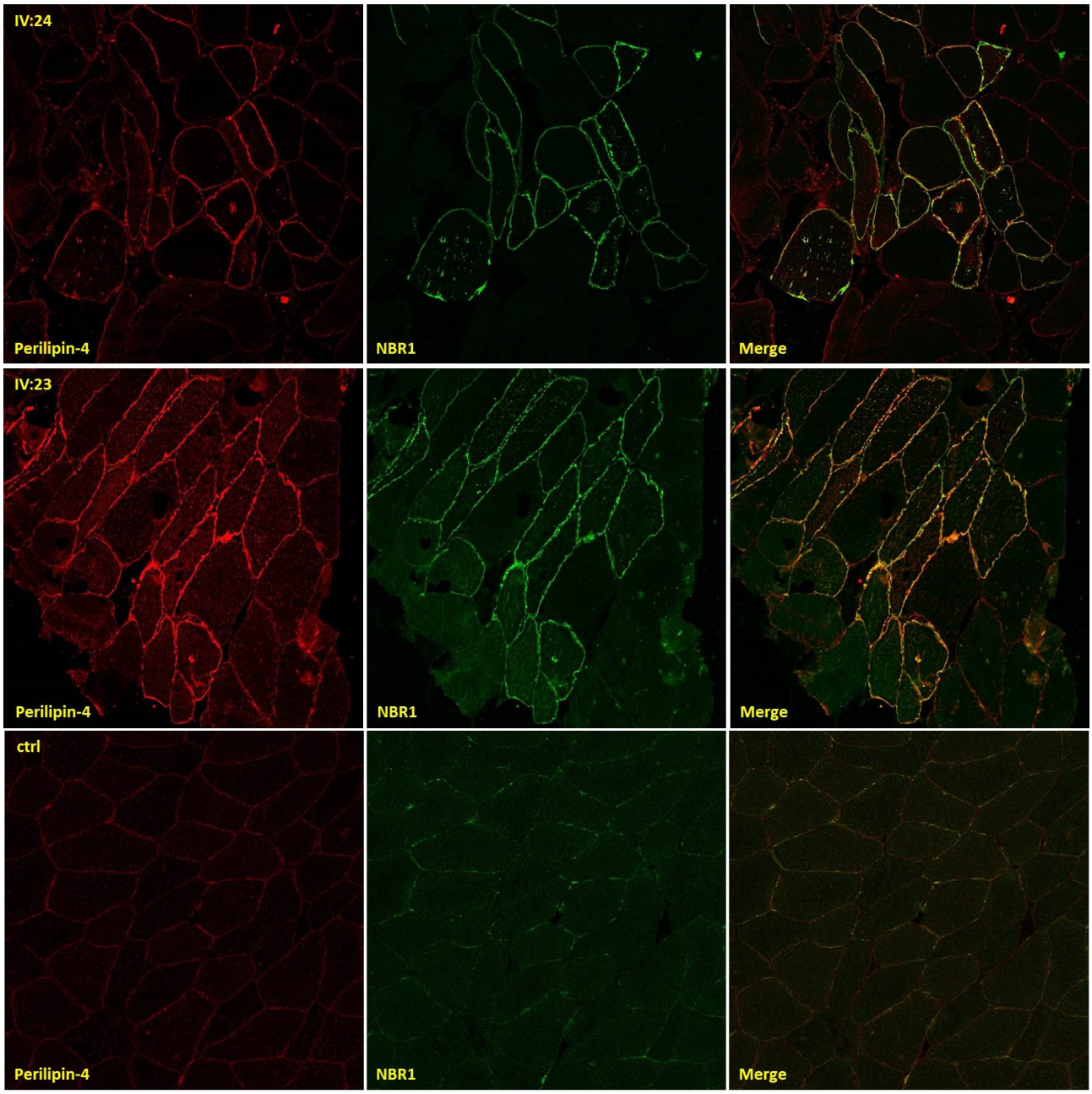


Figure 35. Representative immunohistochemical analysis showing the co-localization between perilipin 4 and NBR1. It is evident the co-localization and the positivity following the phenotype's severity (IV:24 less severe than IV:23). The control muscle biopsy shows no increase of both antibodies' signals.

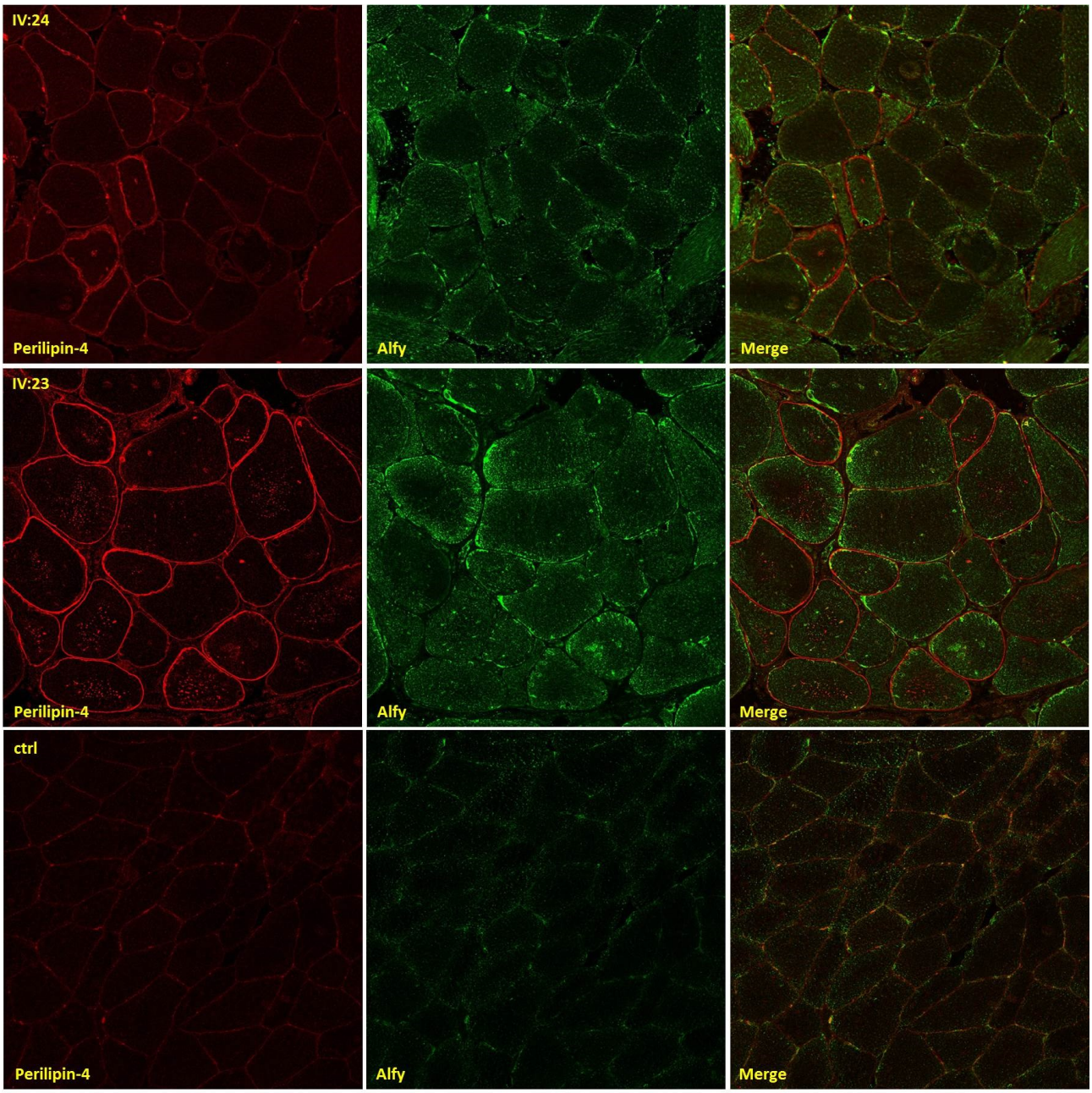


Figure 36. Representative immunohistochemical analysis showing the staining of perilipin 4 and NBR1. A stronger signal for Alfy is observable in the more severely affected patient (bottom panel), without co-localization of the two signals.

4.6 Cell model generation

4.6.1 Plasmids construction

We speculated, based on the patients' muscle biopsies findings, that the pathogenic mechanism leading to this newly identified distal vacuolar myopathy it is likely arising from the misfolding of the mutated isoform due to the coding expansion, followed by the aggregates formation and activation of the aggrephagy pathway and accumulation of vacuoles among the myofibrils.

In order to verify this hypothesis at a molecular level, we switched our attention to a cell model.

Perilipin 4 expression however, despite being present in the mature skeletal muscle tissue, it is not detectable in myoblast cells obtained from the muscle biopsies of the patients, which are proliferative precursor cells able to differentiate and fuse into primary myofibers, necessary for the myogenesis and muscle repair. In fact, in myoblasts, the precursor cells obtained from the skeletal muscle tissue, perilipin 4 is not detectable neither at the RNA level, let alone at the protein level (https://www.proteinatlas.org/ENSG00000167676-PLIN4).

Therefore, we decided to generate by cloning strategies an overexpression model expressing either a wild-type or mutant form of perilipin 4 protein.

We chose to insert the full-length wt and mut cDNAs, retrotranscribed from the RNA extracted from skeletal muscle of a patient, into a mammalian expression vector presenting the fluorescent GFP tag at the N-terminus, thus generating 2 plasmid constructs that we called GFP-PLIN4-FL-WT and GFP-PLIN4-FL-MUT (Fig.37).

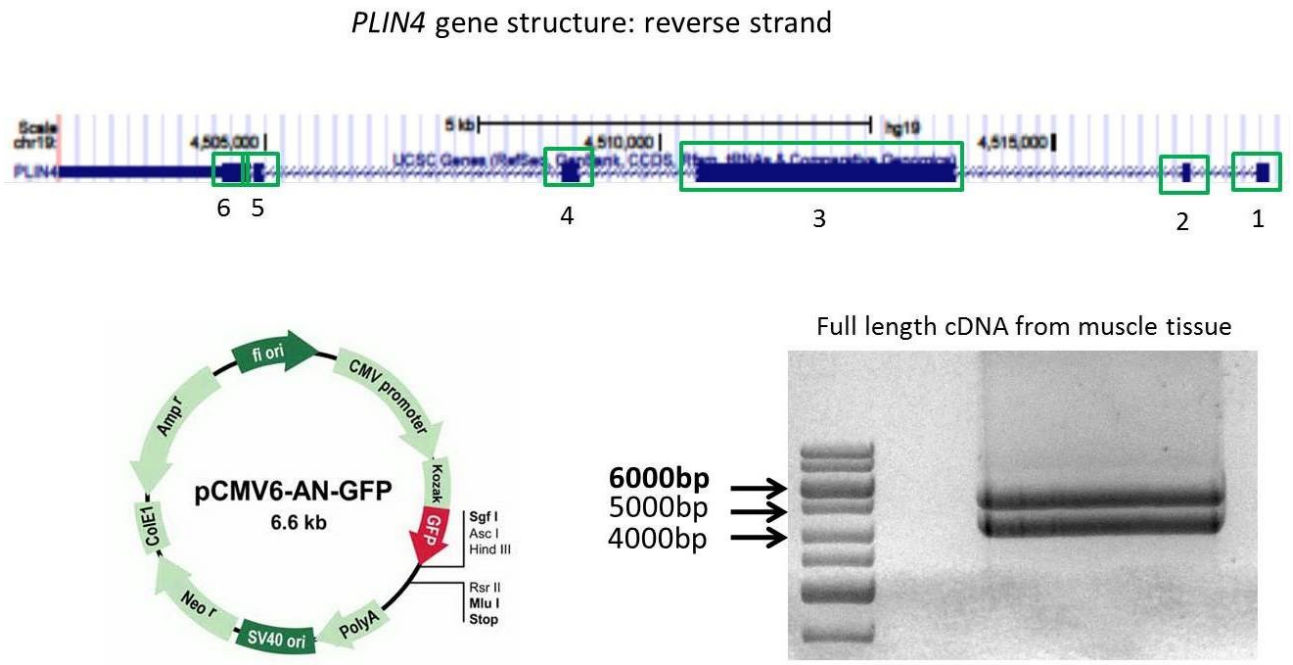


Figure 37. Schematic representation displaying how the plasmid constructs expressing the full-length wild-type or mutated perilipin 4 were designed. Perilipin 4 cDNA obtained from the skeletal muscle was amplified from a patient, each single band was isolated and cloned at the C-term part of the GFP sequence of a commercial plasmid.

Moreover, we also decided to clone the exon 3 alone, encoding for the entire amphipatic helix, in the likelihood that the large size of the entire gene's cDNA, being 4272 nt for the wild-type with extra 891 nt for the mutant allele (for a total of 5163 nt), could become an issue for its exogenous over expression within the host cells. Yet, the size of the exon 3 itself it is rather large, accounting

for 3549 nt for the wild-type and 4350 nt for the mutant sequence. The exon 3 constructs (GFP-PLIN4-EX3-WT and GFP-PLIN4-EX3-MUT) were generated using the same backbone vector as the full-length constructs, but having the GFP expressed at the C-terminal portion of the proteins (Fig.38).

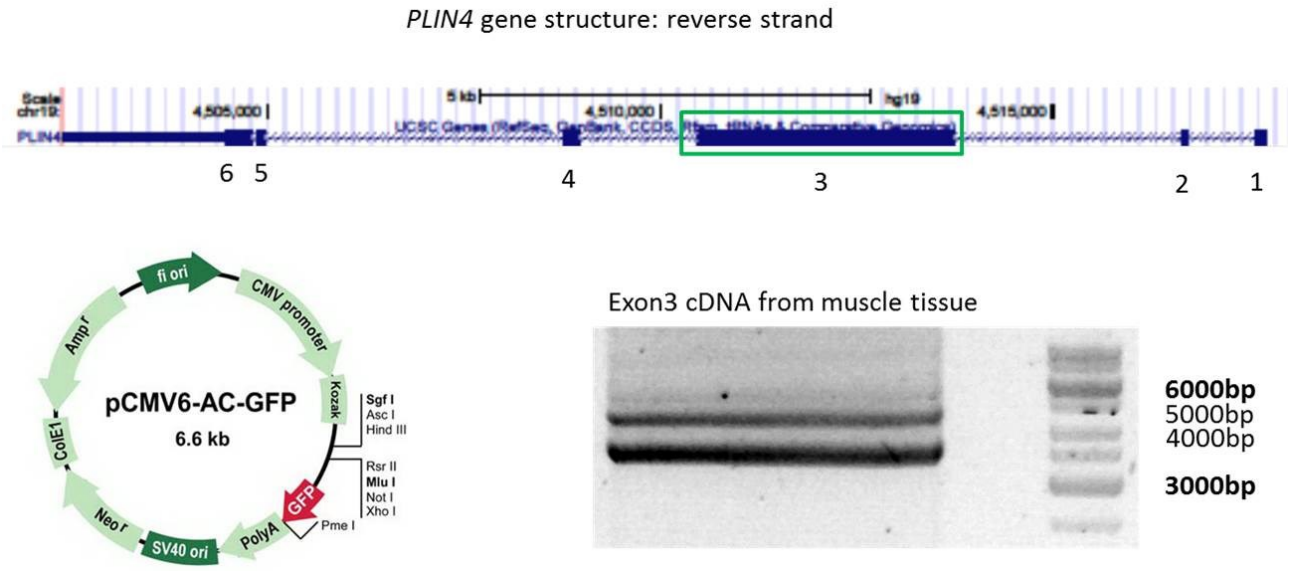


Figure 38. Schematic representation showing the design of the plasmids expressing only the wild-type or mutated exon 3 of the *PLIN4* gene. Genomic DNA was amplified from a patient using the appropriate primers, each single band was isolated and cloned at the N-term part of the GFP sequence of a commercial plasmid.

4.6.2 Western blot validation

We validated our constructs by probing the sizes of the proteins produced by the expression of these plasmids into the host cells.

We transiently transfected HeLa cells with the four constructs and after 48 h of expression we collected the protein extracts for western blot analysis.

The expected protein sizes, theoretically calculated on the number of amino acids in the sequences and including 241 amino acids of the GFP protein, should be 183 kDa and 216 kDa for the GFP-PLIN4-FL-WT and GFP-PLIN4-FL-MUT, respectively, and 153 kDa and 186 kDa for the GFP-PLIN4-EX3-WT and GFP-PLIN4-EX3-MUT, respectively.

The western blot results, using an antibody specific for the GFP protein, revealed, for the full-length proteins, positive bands with sizes consistent with the expected ones. Moreover, probing the same membrane with an antibody recognizing the perilipin 4 protein, we were able to observe the overlap of both signals, by using secondary antibodies conjugated with specific fluorophores.

This confirmed that indeed, the expressed full-length proteins, were the wild-type and mutant perilipin 4 conjugated with the green fluorescent protein.

On the contrary, the GFP-PLIN4-EX3-WT and GFP-PLIN4-EX3-MUT constructs seem to produce proteins that were subjected to a cleavage, an issue that was not prevented even if the necessary proteases inhibitors were added to the cell lysis buffer. This might suggest that a rearrangement or cleavage of the shorter perilipin 4 is occurring within the cells, even before the total protein extract is obtained. So far, we were not able to clarify this issue. A MALDI-TOF mass spectrometry could be useful to try to determine the nature of these smaller peptides. Due to the lack of the perilipin 4 epitope in these shorter perilipin 4 proteins, we were also not able to perform a double staining to verify a correspondence with the GFP antibody signal (Fig.39).

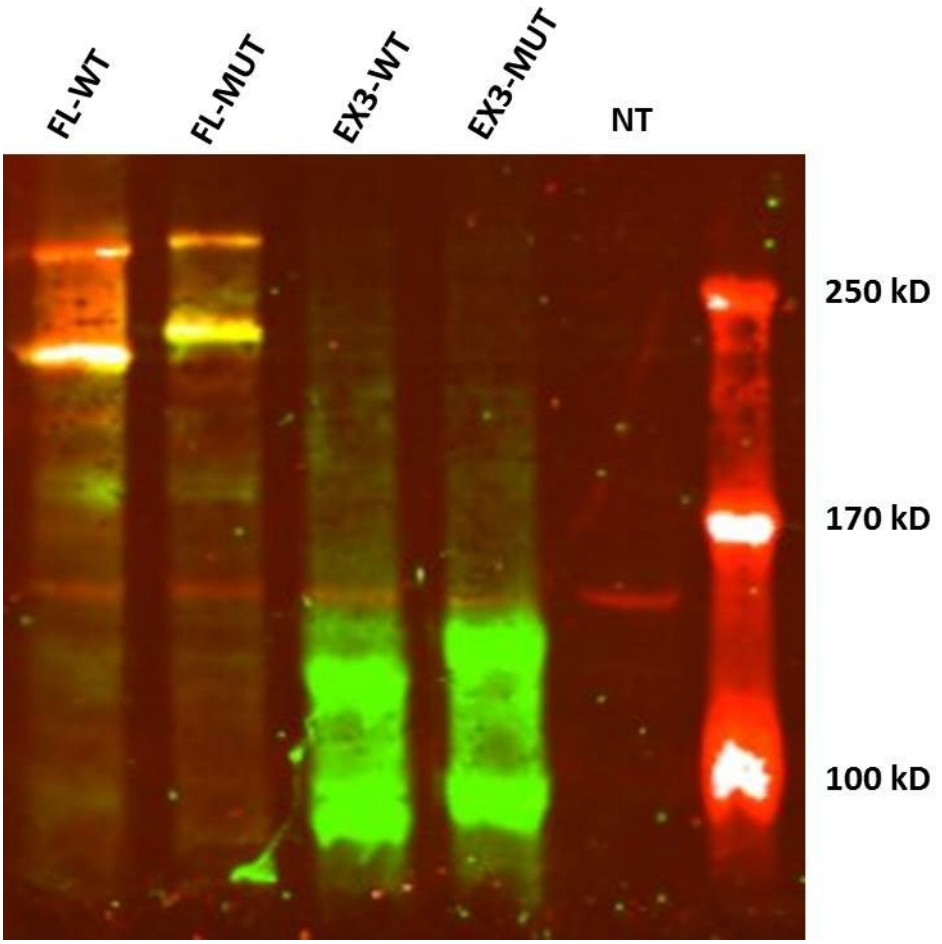


Figure 39. Western blot analysis of total protein extract from transfected HeLa cells. PLIN4 bands detection was obtained with the Odyssey (Li-Cor) system. The membrane was probed with the antibodies recognizing the GFP (signal in green) and perilipin 4 (signal in red). A non-transfected control was included to show specificity of the signals (NT). In the first two lanes are the protein extracts from cells transfected with the full-length plasmids. The sizes of the bands, recognized by both antibodies (green + red) are consistent with the expected molecular weight. The third and fourth lanes are loaded with protein extracts form cells transfected with the exon 3 only plasmid constructs. The antibody against the GFP recognizes bands that are smaller than expected. No

recognition with the perilipin 4 antibody was possible, due to the lack of epitope. An unspecific signal due to perilipin 4 antibody is visible in all lanes including the NT. This signal does not correspond to any of our GFP positive bands.

4.6.3 Immunofluorescence validation

Along with the validation of the sizes of the proteins produced by our plasmid constructs, we also analyzed the correct localization of the exogenous and transiently-expressed perilipin 4, further verifying that the wild-type isoforms were not causing aggregation due to the overexpression.

We transiently transfected HeLa cells with the four plasmids and after 48hr of expression we verified the distribution of the GFP in the fixed cells. The cells were treated at the same time with oleic acid inducing the formation of lipid droplets to better test the behavior of the exogenous proteins.

We observed that at 48hr GFP positive signals surrounding the LDs were visible in all the transfected cells, either expressing the wild-type or the mutant isoforms, both with the full-length (Fig.40) and exon 3-only constructs (Fig.41). Moreover, by staining the full-length transfected cells with the perilipin 4 antibody, we could observe some co-localization, indicating that indeed the GFP positive signals surrounding the LDs are produced by the transiently expressed proteins (Fig.40).

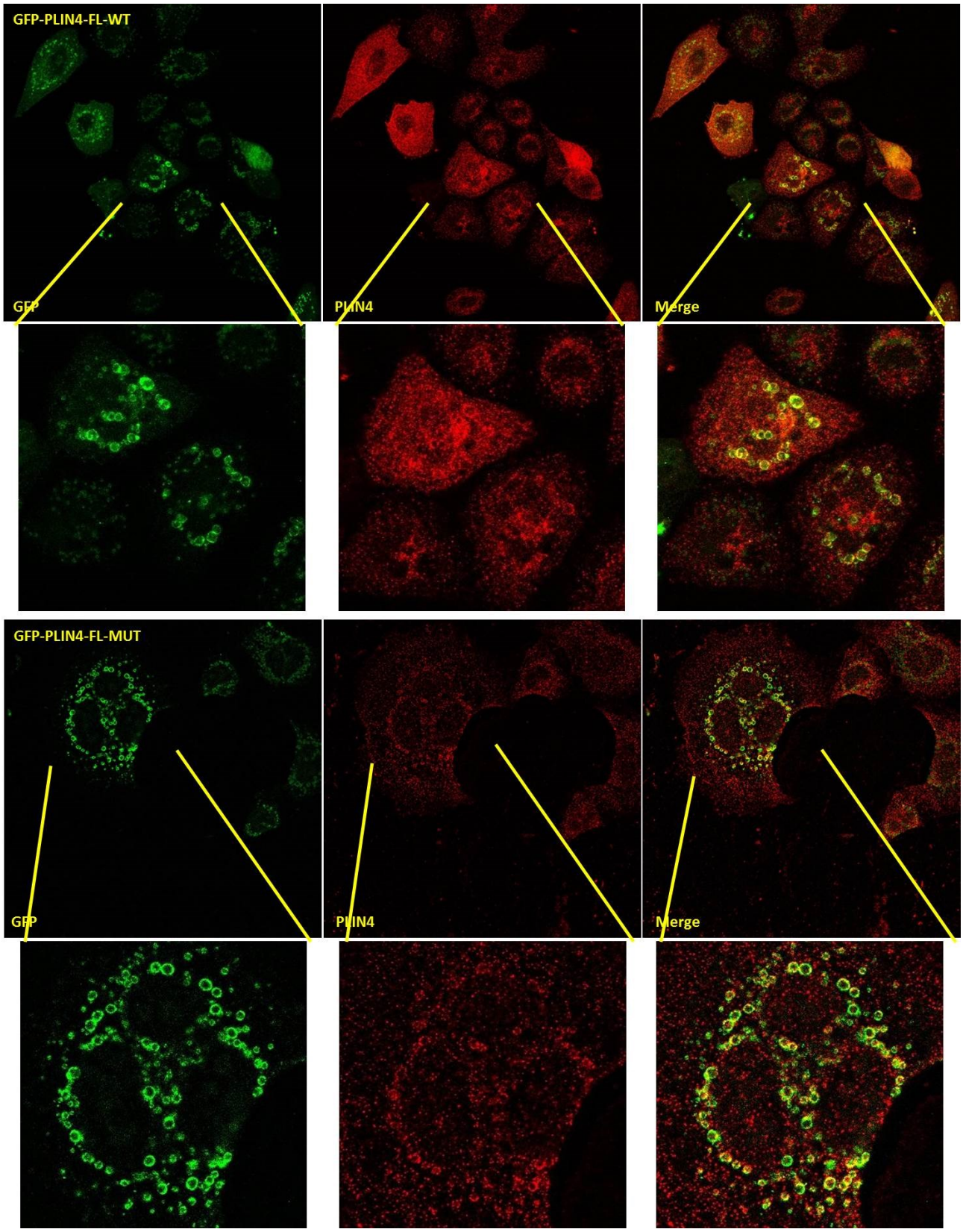


Figure 40. Immunofluorescence images of 48hr transfected cells using the full-length plasmids and treated with 250 µM oleic acid. The top panels show the distribution on the wild-type GFP-perilipin 4 in HeLa cells. In the magnified area, GFP and perilipin 4 positive signals surrounding lipid droplets (not stained) are clearly visible. Moreover, same double positive staining is present in the HeLa cells transfected with the mutant GFP-perilipin 4 (63x).

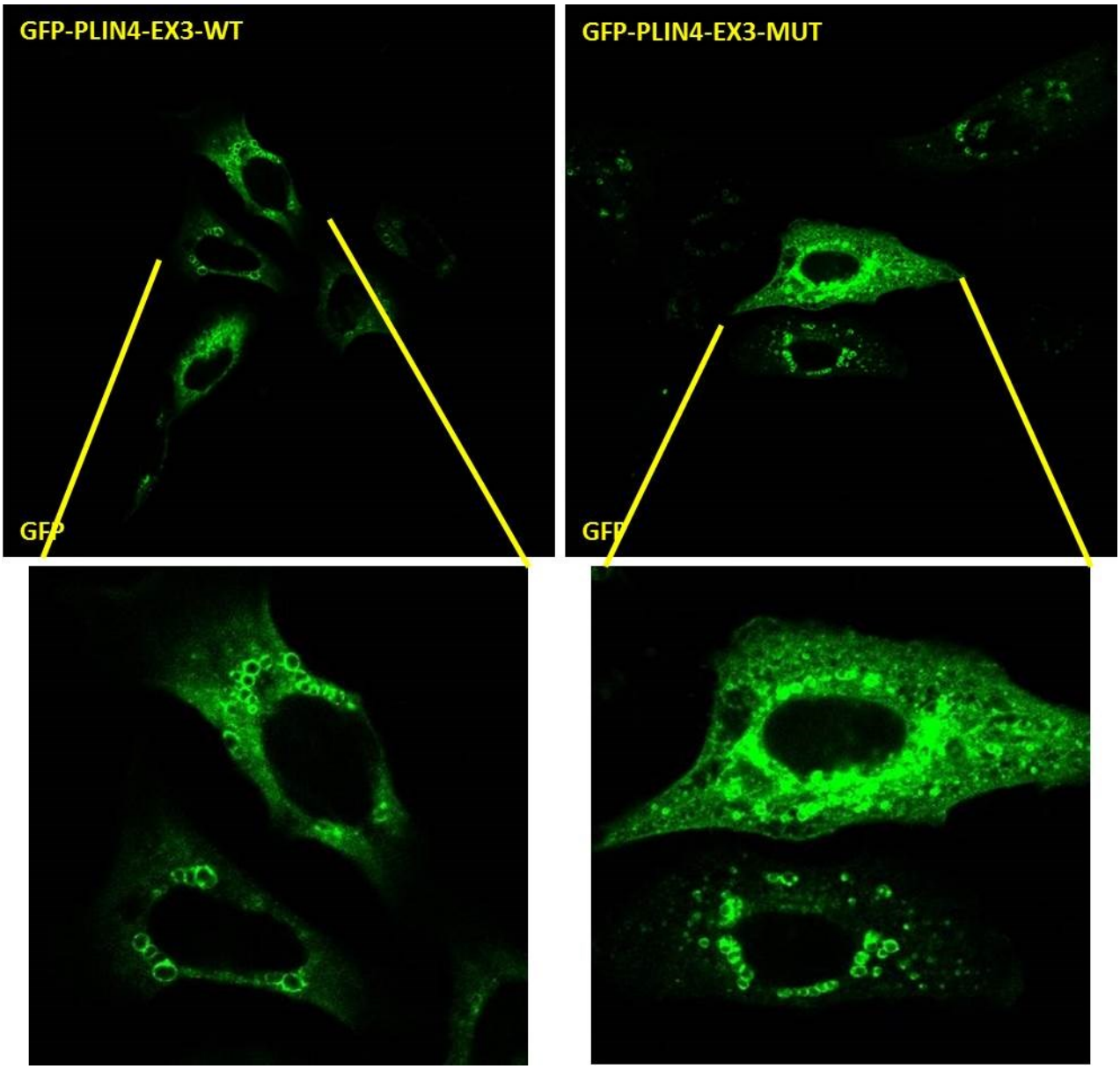


Figure 41. The GFP-PLIN4-EX3-WT (left panel) and GFP-PLIN4-EX3-MUT (right panel) plasmid constructs express perilipin 4-GFP proteins that appear to encircle the LDs present within the cytoplasm of transfected HeLa cells (63x).

4.6.4 Stable cell line generation

Considering that the transient transfection of the mutant isoform did not show the presence of GFP-perilipin 4 aggregates within the cytoplasm, we hypothesize that, because the disease affecting the family is a late onset, a brief expression of the mutant protein might not have been sufficient for causing the aggregation.

In fact, it might be possible that the formation of the aggregates could be caused by a limited efficiency of the clearing system, thus causing the aggregates to be visible only after certain time of mutated perilipin 4 accumulation.

Therefore, we decided to generate stable cell lines that express the protein for a longer time. To this aim, we took advantage of the neomycin resistance expressed by the commercial plasmids used to generate our constructs. We selected the HeLa cells that stably integrated the plasmid within the host genome by adding 400 µM of G418 to the growth medium and allowing the necessary time to remove all the non-integrating cells. This experimental procedure is still ongoing.

We are confident that a stable cell model could be of great help in studying the pathogenic mechanisms leading to this newly identified disorder.

Furthermore, a stable model showing a fluorescent aggregation of the mutant protein could also be used as a screening tool to test commercially available chemical compounds that could help prevent or reduce the formation of perilipin 4 aggregates.

5. Discussion

For many years we have been investigating a large Italian family affected by an autosomal dominant distal vacuolar myopathy. The onset of the disorder is around the third or fourth decade of life with a certain degree of variability. The weakness initially affects the distal muscles but it extends, with a slow progression, to all the muscle districts with time. No other tissues beside the skeletal muscle seem to be affected, including the cardiac muscle which doesn't show at the clinical level, signs of impairment. However, no autoptic samples were ever analyzed so we cannot exclude that a subclinical alteration is anyway present in this tissue.

The histology of the skeletal muscle shows the presence of numerous autophagic vacuoles that stain positive for p62/SQSTM1 and for the ubiquitinated proteins (FK2). Within the muscle the continuous formation and fusion of vacuoles engulf the fibers, disrupting their fine organization and altering their contractile abilities, thus resulting in muscle fiber atrophy.

Genetic analysis performed using the latest Next-Generation Sequencing techniques available, did not help reveal the causative gene whose locus was identified in a 5.12 Mb region of chromosome 19 (19p13.3).

Only by using a reverse approach, first identifying the proteomic content of the vacuoles, and then applying a long-range Oxford Nanopore Technology sequencing approach, we were able to determine the pathogenic mutation. This consists in a large coding expansion of 9 x 33mer, resulting in 297 extra amino acids within the amphipathic helix of perilipin 4, encoded by the *PLIN4* gene, located in the linked region [Čopič et al, 2018]. Perilipin 4 is involved in the

metabolism of Lipid Droplets [Guo et al, 2009] and is mainly expressed in the white adipose tissue and to a lesser extent in the skeletal muscle [Pourteymour et al, 2015].

Remarkably, the amphipathic helix of perilipin 4 is structurally similar to the helix of α-synuclein and exchangeable apolipoprotein A1, proteins that are known to interact with lipids [Jao et al, 2008; Frank & Marcel, 2000] and to be prone to aggregation and precipitation under disease conditions [Alegre-Abarrategui et al, 2019; Mehra et al, 2019; Das et al, 2014; Arciello et al, 2016]. Interestingly, all known mutations (all missense) of α-synuclein and ApoA1 in familial Parkinson disease and amyloidosis, respectively, are located in the amphipathic domains in these proteins.

We speculated that the expansion in the coding region of *PLIN4* gene pathogenically elongates its protein product causing it to misfold and precipitate. These precipitates will deposit as aggresomes and are cleared through ubiquitin mediated autophagy (aggrephagy). Specifically, the p62/SQSTM1 protein, along with several other components, will scaffold the aggresome ubiquitins with LC3, which causes the aggregate to become enclosed in a lipid bilayer membrane (autophagosome) that merges with lysosomes for autophagic degradation [Knaevelsrud et al, 2010]. As part of the clearance of ubiquitinated protein aggregates, p62/SQSTM1 interacts with NBR1, and the two, as an autophagy receptor complex and through their shared LC3-Interacting Regions (LIRs), bridge the ubiquitinated misfolded proteins with LC3 [Kirkin et al, 2009a]. Meanwhile, WDFY3 shuttles from the nucleus to the cytoplasm to scaffold the overall structure with PtdIns3P-containing membranes and promote encapsulation of the aggregates in autophagosomes [Clausen et al, 2010].

Most likely, an imbalance between formation and removal of aggregates, eventually leads to the accumulation of vacuoles.

To better understand the mechanisms leading to this disorder we wanted to investigate at the molecular level the involvement of the aggrephagy pathway and the formation and clearance of the aggregates. We therefore focused our attention on myoblasts derived from the muscle biopsy of patients, hoping to be able to visualize the formation of aggregates within the cell cytoplasm. Unfortunately, no perilipin 4 expression was detectable in this cell type, due to the fact that myoblasts are precursors of muscle fibers, still immature and therefore not expressing all the proteins that are expressed in mature fibers.

Therefore, we decided to develop an overexpression cellular model that could be used to dissect all the pathogenic aspects involved and could become a valuable screening tool to test

commercially available compounds that might act in preventing or reducing the formation of the aggregates.

The first attempts to generate this cellular model, based on a transient transfection, did not fully succeeded, possibly due to the short time of expression of the mutant protein. To overcome this limitation, we are generating clones that stably overexpress the mutated isoform of perilipin 4. Moreover, different cell types and different cell treatments could help generate a cell model that could better recapitulate what we observe in patients' skeletal muscles. For instance the use of C1C12 mouse myoblasts could better represent the energy requirement of the muscle tissue, therefore resulting in a more appropriate model.

In addition, we are planning to differentiate the mesenchimal stem cells (MSC) derived from the patients' iPS into adipocytes, a cell type in which perilipin 4 is constitutively expressed. We hope that we will be able to observe the presence of perilipin 4 aggregates in the patients' adipocytes, and then use this model to investigate all the many unresolved questions that still baffle us.

Another strategy that we are planning to use is the generation of an animal model, particularly using the small freshwater fish zebrafish (*Danio rerio*), a model that has already been exploited to recreate muscle disorders [Widrick et al, 2016] and which is particularly suitable for the identification of compounds with possible therapeutic effects [Kawahara et al, 2011].

Activation of aggrephagy appears to underlie other myopathies, as indicated by reports of upregulation of some aggrephagy proteins in other form of skeletal muscle disease. Notably, an altered deposition of the NBR1 protein and involvement of the aggresome [D'Agostino et al, 2011; Sabatelli et al, 2014] has been reported in sporadic inclusion body myositis (sIBM) and in muscle disorders due to mutations in the *FHL1* gene. Both conditions are characterized by the presence of vacuoles within muscle fibers. Furthermore, accumulation or aggregation of misfolded proteins, which usually consist of cytoskeletal elements or their associated proteins has been found in muscle fibers in myofibrillar myopathies. Accumulation of misfolded protein aggregates in skeletal muscle occurs also with proteotoxic gene mutations in neurodegenerative diseases such as amyotrophic lateral sclerosis [Dobrowolny et al, 2009] and Huntington's disease [Tomioka et al, 2017].

This is the first report of an amphipathic domain repeat expansion in a human disease, and identification of the expansion was only possible with a long-read sequencing approach and the first report ever published of a disorder due to perilipin 4 alteration. The perilipin 4-related

myopathy appears to be a further disease related to the formation of proteotoxic aggregates in muscle fibers.

6. References

- Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002;30(1):97-101. doi:10.1038/ng786
- Alegre-Abarrategui J, Brimblecombe KR, Roberts RF, et al. Selective vulnerability in α-synucleinopathies. Acta Neuropathol. 2019;138(5):681-704. doi:10.1007/s00401-019-02010-2
- Amerik AY, Hochstrasser M. Mechanism and function of deubiquitinating enzymes. Biochim Biophys Acta. 2004;1695(1-3):189-207. doi:10.1016/j.bbamcr.2004.10.003
- Arciello A, Piccoli R, Monti DM. Apolipoprotein A-I: the dual face of a protein. FEBS Lett. 2016;590(23):4171-4179. doi:10.1002/1873-3468.12468
- Bickel PE, Tansey JT, Welte MA. PAT proteins, an ancient family of lipid droplet proteins that regulate cellular lipid stores. Biochim Biophys Acta. 2009;1791(6):419-440. doi:10.1016/j.bbalip.2009.04.002
- Bindesbøll C, Berg O, Arntsen B, Nebb HI, Dalen KT. Fatty acids regulate perilipin5 in muscle by activating PPARδ. J Lipid Res. 2013;54(7):1949-1963. doi:10.1194/jlr.M038992
- Bjørkøy G, Lamark T, Brech A, et al. p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. J Cell Biol. 2005;171(4):603-614. doi:10.1083/jcb.200507002
- Bosma M, Minnaard R, Sparks LM, et al. The lipid droplet coat protein perilipin 5 also localizes to muscle mitochondria. Histochem Cell Biol. 2012;137(2):205-216. doi:10.1007/s00418-011-0888-x
- Bosma M. Lipid droplet dynamics in skeletal muscle. Exp Cell Res. 2016;340(2):180-186. doi:10.1016/j.yexcr.2015.10.023
- Brasaemle DL. Thematic review series: adipocyte biology. The perilipin family of structural lipid droplet proteins: stabilization of lipid droplets and control of lipolysis. J Lipid Res. 2007;48(12):2547-2559. doi:10.1194/jlr.R700014-JLR200
- Budenholzer L, Cheng CL, Li Y, Hochstrasser M. Proteasome Structure and Assembly. J Mol Biol. 2017;429(22):3500-3524. doi:10.1016/j.jmb.2017.05.027
- Bussell R Jr, Eliezer D. A structural and functional role for 11-mer repeats in alpha-synuclein and other exchangeable lipid binding proteins. J Mol Biol. 2003;329(4):763-778. doi:10.1016/s0022-2836(03)00520-5
- Cavey JR, Ralston SH, Hocking LJ, et al. Loss of ubiquitin-binding associated with Paget's disease of bone p62 (SQSTM1) mutations. J Bone Miner Res. 2005;20(4):619-624. doi:10.1359/JBMR.041205
- Chen Y, Klionsky DJ. The regulation of autophagy - unanswered questions. J Cell Sci. 2011;124(Pt 2):161-170. doi:10.1242/jcs.064576
- Ciechanover A, Iwai K. The ubiquitin system: from basic mechanisms to the patient bed. IUBMB Life. 2004;56(4):193-201. doi:10.1080/1521654042000223616
- Ciechanover A. Proteolysis: from the lysosome to ubiquitin and the proteasome. Nat Rev Mol Cell Biol. 2005;6(1):79-87. doi:10.1038/nrm1552

- Ciechanover A, Kwon YT. Protein Quality Control by Molecular Chaperones in Neurodegeneration. Front Neurosci. 2017;11:185. Published 2017 Apr 6. doi:10.3389/fnins.2017.00185
- Clague MJ, Urbé S. Ubiquitin: same molecule, different degradation pathways. Cell. 2010;143(5):682-685. doi:10.1016/j.cell.2010.11.012
- Clausen TH, Lamark T, Isakson P, et al. p62/SQSTM1 and ALFY interact to facilitate the formation of p62 bodies/ALIS and their degradation by autophagy. Autophagy. 2010;6(3):330-344. doi:10.4161/auto.6.3.11226
- Čopič A, Antoine-Bally S, Giménez-Andrés M, et al. A giant amphipathic helix from a perilipin that is adapted for coating lipid droplets. Nat Commun. 2018;9(1):1332. Published 2018 Apr 6. doi:10.1038/s41467-018-03717-8
- D'Agostino C, Nogalska A, Cacciottolo M, Engel WK, Askanas V. Abnormalities of NBR1, a novel autophagy-associated protein, in muscle fibers of sporadic inclusion-body myositis. Acta Neuropathol. 2011;122(5):627-636. doi:10.1007/s00401-011-0874-3
- Das M, Mei X, Jayaraman S, Atkinson D, Gursky O. Amyloidogenic mutations in human apolipoprotein A-I are not necessarily destabilizing - a common mechanism of apolipoprotein A-I misfolding in familial amyloidosis and atherosclerosis. FEBS J. 2014;281(11):2525-2542. doi:10.1111/febs.12809
- Di Blasi C, Moghadaszadeh B, Ciano C, et al. Abnormal lysosomal and ubiquitin-proteasome pathways in 19p13.3 distal myopathy. Ann Neurol. 2004;56(1):133-138. doi:10.1002/ana.20158
- Dice JF. Peptide sequences that target cytosolic proteins for lysosomal proteolysis. Trends Biochem Sci. 1990;15(8):305-309. doi:10.1016/0968-0004(90)90019-8
- Dobrowolny G, Aucello M, Rizzuto E, et al. Skeletal muscle is a primary target of SOD1G93A-mediated toxicity [published correction appears in Cell Metab. 2009 Jan;9(1):110. Bonconpagni, Simona [corrected to Boncompagni, Simona]]. Cell Metab. 2008;8(5):425-436. doi:10.1016/j.cmet.2008.09.002
- Dobson CM. Protein folding and misfolding. Nature. 2003;426(6968):884-890. doi:10.1038/nature02261
- Donaldson KM, Li W, Ching KA, Batalov S, Tsai CC, Joazeiro CA. Ubiquitin-mediated sequestration of normal cellular proteins into polyglutamine aggregates. Proc Natl Acad Sci U S A. 2003;100(15):8892-8897. doi:10.1073/pnas.1530212100
- Felice KJ. Distal Myopathies. Neurol Clin. 2020;38(3):637-659. doi:10.1016/j.ncl.2020.03.007
- Filimonenko M, Isakson P, Finley KD, et al. The selective macroautophagic degradation of aggregated proteins requires the PI3P-binding protein Alfy. Mol Cell. 2010;38(2):265-279. doi:10.1016/j.molcel.2010.04.007
- Fink AL. Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold Des. 1998;3(1):R9-R23. doi:10.1016/S1359-0278(98)00002-9
- Fischer D, Kley RA, Strach K, et al. Distinct muscle imaging patterns in myofibrillar myopathies. Neurology. 2008;71(10):758-765. doi:10.1212/01.wnl.0000324927.28817.9b
- Frank PG, Marcel YL. Apolipoprotein A-I: structure-function relationships. J Lipid Res. 2000;41(6):853-872.
- Galluzzi L, Baehrecke EH, Ballabio A, et al. Molecular definitions of autophagy and related processes. EMBO J. 2017;36(13):1811-1836. doi:10.15252/embj.201796697

- Gandotra S, Le Dour C, Bottomley W, et al. Perilipin deficiency and autosomal dominant partial lipodystrophy. N Engl J Med. 2011;364(8):740-748. doi:10.1056/NEJMoa1007487
- Garcia-Mata R, Gao YS, Sztul E. Hassles with taking out the garbage: aggravating aggresomes. Traffic. 2002;3(6):388-396. doi:10.1034/j.1600-0854.2002.30602.x
- Giménez-Andrés M, Čopič A, Antonny B. The Many Faces of Amphipathic Helices. Biomolecules. 2018;8(3):45. Published 2018 Jul 5. doi:10.3390/biom8030045
- Giménez-Andrés M, Emeršič T, Antoine-Bally S, et al. Exceptional stability of a perilipin on lipid droplets depends on its polar residues, suggesting multimeric assembly. Elife. 2021;10:e61401. Published 2021 Apr 15. doi:10.7554/eLife.61401
- Glickman MH, Ciechanover A. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. Physiol Rev. 2002;82(2):373-428. doi:10.1152/physrev.00027.2001
- Gowers WR. A Lecture on Myopathy and a Distal Form: Delivered at the National Hospital for the Paralysed and Epileptic. Br Med J. 1902;2(2167):89-92. doi:10.1136/bmj.2.2167.89
- Granneman JG, Kimler VA, Zhang H, et al. Lipid droplet biology and evolution illuminated by the characterization of a novel perilipin in teleost fish. Elife. 2017;6:e21771. Published 2017 Feb 28. doi:10.7554/eLife.21771
- Guo Y, Cordes KR, Farese RV Jr, Walther TC. Lipid droplets at a glance. J Cell Sci. 2009;122(Pt 6):749-752. doi:10.1242/jcs.037630
- Hickenbottom SJ, Kimmel AR, Londos C, Hurley JH. Structure of a lipid droplet protein; the PAT family member TIP47. Structure. 2004;12(7):1199-1207. doi:10.1016/j.str.2004.04.021
- Hsieh K, Lee YK, Londos C, Raaka BM, Dalen KT, Kimmel AR. Perilipin family members preferentially sequester to either triacylglycerol-specific or cholesteryl-ester-specific intracellular lipid storage droplets. J Cell Sci. 2012;125(Pt 17):4067-4076. doi:10.1242/jcs.104943
- Huang X, Luan B, Wu J, Shi Y. An atomic structure of the human 26S proteasome. Nat Struct Mol Biol. 2016;23(9):778-785. doi:10.1038/nsmb.3273
- Ichimura Y, Kumanomidou T, Sou YS, et al. Structural basis for sorting mechanism of p62 in selective autophagy. J Biol Chem. 2008;283(33):22847-22857. doi:10.1074/jbc.M802182200
- Isakson P, Holland P, Simonsen A. The role of ALFY in selective autophagy. Cell Death Differ. 2013;20(1):12-20. doi:10.1038/cdd.2012.66
- Isogai S, Morimoto D, Arita K, et al. Crystal structure of the ubiquitin-associated (UBA) domain of p62 and its interaction with ubiquitin. J Biol Chem. 2011;286(36):31864-31874. doi:10.1074/jbc.M111.259630
- Itabe H, Yamaguchi T, Nimura S, Sasabe N. Perilipins: a diversity of intracellular lipid droplet proteins. Lipids Health Dis. 2017;16(1):83. Published 2017 Apr 28. doi:10.1186/s12944-017-0473-y
- Iwata A, Riley BE, Johnston JA, Kopito RR. HDAC6 and microtubules are required for autophagic degradation of aggregated huntingtin. J Biol Chem. 2005;280(48):40282-40292. doi:10.1074/jbc.M508786200
- Jao CC, Hegde BG, Chen J, Haworth IS, Langen R. Structure of membrane-bound alpha-synuclein from site-directed spin labeling and computational refinement. Proc Natl Acad Sci U S A. 2008;105(50):19666-19671. doi:10.1073/pnas.0807826105

- Jung CH, Jun CB, Ro SH, et al. ULK-Atg13-FIP200 complexes mediate mTOR signaling to the autophagy machinery. Mol Biol Cell. 2009;20(7):1992-2003. doi:10.1091/mbc.e08-12-1249
- Kaushik S, Cuervo AM. Degradation of lipid droplet-associated proteins by chaperone-mediated autophagy facilitates lipolysis. Nat Cell Biol. 2015;17(6):759-770. doi:10.1038/ncb3166
- Kaushik S, Cuervo AM. The coming of age of chaperone-mediated autophagy. Nat Rev Mol Cell Biol. 2018;19(6):365-381. doi:10.1038/s41580-018-0001-6
- Kawaguchi Y, Kovacs JJ, McLaurin A, Vance JM, Ito A, Yao TP. The deacetylase HDAC6 regulates aggresome formation and cell viability in response to misfolded protein stress. Cell. 2003;115(6):727-738. doi:10.1016/s0092-8674(03)00939-5
- Kawahara G, Karpf JA, Myers JA, Alexander MS, Guyon JR, Kunkel LM. Drug screening in a zebrafish model of Duchenne muscular dystrophy. Proc Natl Acad Sci U S A. 2011;108(13):5331-5336. doi:10.1073/pnas.1102116108
- Kimmel AR, Brasaemle DL, McAndrews-Hill M, Sztalryd C, Londos C. Adoption of PERILIPIN as a unifying nomenclature for the mammalian PAT-family of intracellular lipid storage droplet proteins. J Lipid Res. 2010;51(3):468-471. doi:10.1194/jlr.R000034
- Kirkin V, Lamark T, Sou YS, et al. A role for NBR1 in autophagosomal degradation of ubiquitinated substrates. Mol Cell. 2009;33(4):505-516. doi:10.1016/j.molcel.2009.01.020 a
- Kirkin V, Lamark T, Johansen T, Dikic I. NBR1 cooperates with p62 in selective autophagy of ubiquitinated targets. Autophagy. 2009;5(5):732-733. doi:10.4161/auto.5.5.8566 b
- Klionsky DJ, Eskelinen EL, Deretic V. Autophagosomes, phagosomes, autolysosomes, phagolysosomes, autophagolysosomes... wait, I'm confused. Autophagy. 2014;10(4):549-551. doi:10.4161/auto.28448
- Knaevelsrud H, Simonsen A. Fighting disease by selective autophagy of aggregate-prone proteins. FEBS Lett. 2010;584(12):2635-2645. doi:10.1016/j.febslet.2010.04.041
- Kopito RR. Aggresomes, inclusion bodies and protein aggregation. Trends Cell Biol. 2000;10(12):524-530. doi:10.1016/s0962-8924(00)01852-3
- Kopito RR. The missing linker: an unexpected role for a histone deacetylase. Mol Cell. 2003;12(6):1349-1351. doi:10.1016/s1097-2765(03)00498-2
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722-736. doi:10.1101/gr.215087.116
- Kuusisto E, Salminen A, Alafuzoff I. Ubiquitin-binding protein p62 is present in neuronal and glial inclusions in human tauopathies and synucleinopathies. Neuroreport. 2001;12(10):2085-2090. doi:10.1097/00001756-200107200-00009
- Kuusisto E, Salminen A, Alafuzoff I. Early accumulation of p62 in neurofibrillary tangles in Alzheimer's disease: possible role in tangle formation. Neuropathol Appl Neurobiol. 2002;28(3):228-237. doi:10.1046/j.1365-2990.2002.00394.x
- Lamark T, Johansen T. Aggrephagy: selective disposal of protein aggregates by macroautophagy. Int J Cell Biol. 2012;2012:736905. doi:10.1155/2012/736905
- Lee JY, Koga H, Kawaguchi Y, et al. HDAC6 controls autophagosome maturation essential for ubiquitin-selective quality-control autophagy. EMBO J. 2010;29(5):969-980. doi:10.1038/emboj.2009.405

- Li Y, Lee S, Langleite T, et al. Subsarcolemmal lipid droplet responses to a combined endurance and strength exercise intervention. Physiol Rep. 2014;2(11):e12187. Published 2014 Nov 20. doi:10.14814/phy2.12187
- Liu J, Aoki M, Illa I, et al. Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. Nat Genet. 1998;20(1):31-36. doi:10.1038/1682
- Long J, Garner TP, Pandya MJ, et al. Dimerisation of the UBA domain of p62 inhibits ubiquitin binding and regulates NF-kappaB signalling. J Mol Biol. 2010;396(1):178-194. doi:10.1016/j.jmb.2009.11.032
- Lystad AH, Ichimura Y, Takagi K, et al. Structural determinants in GABARAP required for the selective binding and recruitment of ALFY to LC3B-positive structures. EMBO Rep. 2014;15(5):557-565. doi:10.1002/embr.201338003
- Maerkens A, Olivé M, Schreiner A, et al. New insights into the protein aggregation pathology in myotilinopathy by combined proteomic and immunolocalization analyses. Acta Neuropathol Commun. 2016;4:8. Published 2016 Feb 3. doi:10.1186/s40478-016-0280-0
- Malenfant P, Joanisse DR, Thériault R, Goodpaster BH, Kelley DE, Simoneau JA. Fat content in individual muscle fibers of lean and obese subjects. Int J Obes Relat Metab Disord. 2001;25(9):1316-1321. doi:10.1038/sj.ijo.0801733
- Markossian KA, Kurganov BI. Protein folding, misfolding, and aggregation. Formation of inclusion bodies and aggresomes. Biochemistry (Mosc). 2004;69(9):971-984. doi:10.1023/b:biry.0000043539.07961.4c
- Martin S, Parton RG. Caveolin, cholesterol, and lipid bodies. Semin Cell Dev Biol. 2005;16(2):163-174. doi:10.1016/j.semcdb.2005.01.007
- Martinez-Botas J, Anderson JB, Tessier D, et al. Absence of perilipin results in leanness and reverses obesity in Lepr(db/db) mice. Nat Genet. 2000;26(4):474-479. doi:10.1038/82630
- Matsumoto G, Wada K, Okuno M, Kurosawa M, Nukina N. Serine 403 phosphorylation of p62/SQSTM1 regulates selective autophagic clearance of ubiquitinated proteins. Mol Cell. 2011;44(2):279-289. doi:10.1016/j.molcel.2011.07.039
- Mehra S, Sahay S, Maji SK. α-Synuclein misfolding and aggregation: Implications in Parkinson's disease pathogenesis. Biochim Biophys Acta Proteins Proteom. 2019;1867(10):890-908. doi:10.1016/j.bbapap.2019.03.001
- Milhorat AT, Wolff HG. Studies in diseases of muscle: XIII. Progressive muscular dystrophy of atrophic distal type; report on a family; report of autopsy. Arch NeurPsych. 1943;49(5):655–664. doi:10.1001/archneurpsyc.1943.02290170025002
- Miyoshi K, Kawai H, Iwasa M, Kusaka K, Nishino H. Autosomal recessive distal muscular dystrophy as a new type of progressive muscular dystrophy. Seventeen cases in eight families including an autopsied case. Brain. 1986;109 (Pt 1):31-54. doi:10.1093/brain/109.1.31
- Morales PE, Bucarey JL, Espinosa A. Muscle Lipid Metabolism: Role of Lipid Droplets and Perilipins. J Diabetes Res. 2017;2017:1789395. doi:10.1155/2017/1789395
- Nonaka I, Sunohara N, Ishiura S, Satoyoshi E. Familial distal myopathy with rimmed vacuole and lamellar (myeloid) body formation. J Neurol Sci. 1981;51(1):141-155. doi:10.1016/0022-510x(81)90067-8

- Odagiri S, Tanji K, Mori F, Kakita A, Takahashi H, Wakabayashi K. Autophagic adapter protein NBR1 is localized in Lewy bodies and glial cytoplasmic inclusions and is involved in aggregate formation in α-synucleinopathy. Acta Neuropathol. 2012;124(2):173-186. doi:10.1007/s00401-012-0975-7
- Ohsaki Y, Maeda T, Maeda M, Tauchi-Sato K, Fujimoto T. Recruitment of TIP47 to lipid droplets is controlled by the putative hydrophobic cleft. Biochem Biophys Res Commun. 2006;347(1):279-287. doi:10.1016/j.bbrc.2006.06.074
- Palmio J, Jonson PH, Inoue M, et al. Mutations in the J domain of DNAJB6 cause dominant distal myopathy. Neuromuscul Disord. 2020;30(1):38-46. doi:10.1016/j.nmd.2019.11.005
- Pankiv S, Clausen TH, Lamark T, et al. p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. J Biol Chem. 2007;282(33):24131-24145. doi:10.1074/jbc.M702824200
- Parzych KR, Klionsky DJ. An overview of autophagy: morphology, mechanism, and regulation. Antioxid Redox Signal. 2014;20(3):460-473. doi:10.1089/ars.2013.5371
- Pickart CM, Eddins MJ. Ubiquitin: structures, functions, mechanisms. Biochim Biophys Acta. 2004;1695(1-3):55-72. doi:10.1016/j.bbamcr.2004.09.019
- Ploegh HL. A lipid-based model for the creation of an escape hatch from the endoplasmic reticulum. Nature. 2007;448(7152):435-438. doi:10.1038/nature06004
- Pourteymour S, Lee S, Langleite TM, et al. Perilipin 4 in human skeletal muscle: localization and effect of physical activity. Physiol Rep. 2015;3(8):e12481. doi:10.14814/phy2.12481
- Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-575. doi:10.1086/519795
- Reinstein E, Ciechanover A. Narrative review: protein degradation and human diseases: the ubiquitin connection. Ann Intern Med. 2006;145(9):676-684. doi:10.7326/0003-4819-145-9-200611070-00010
- Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-26. doi:10.1038/nbt.1754
- Sabatelli P, Castagnaro S, Tagliavini F, et al. Aggresome-Autophagy Involvement in a Sarcopenic Patient with Rigid Spine Syndrome and a p.C150R Mutation in FHL1 Gene. Front Aging Neurosci. 2014;6:215. Published 2014 Aug 19. doi:10.3389/fnagi.2014.00215
- Savarese M, Palmio J, Poza JJ, et al. Actininopathy: A new muscular dystrophy caused by ACTN2 dominant mutations. Ann Neurol. 2019;85(6):899-906. doi:10.1002/ana.25470
- Savarese M, Sarparanta J, Vihola A, et al. Panorama of the distal myopathies. Acta Myol. 2020;39(4):245-265. Published 2020 Dec 1. doi:10.36185/2532-1900-028.
- Schuck S. Microautophagy - distinct molecular mechanisms handle cargoes of many sizes. J Cell Sci. 2020;133(17):jcs246322. Published 2020 Sep 9. doi:10.1242/jcs.246322
- Segrest JP, Jones MK, De Loof H, Brouillette CG, Venkatachalapathi YV, Anantharamaiah GM. The amphipathic helix in the exchangeable apolipoproteins: a review of secondary structure and function. J Lipid Res. 1992;33(2):141-166.
- Settembre C, Zoncu R, Medina DL, et al. A lysosome-to-nucleus signalling mechanism senses and regulates the lysosome via mTOR and TFEB. EMBO J. 2012;31(5):1095-1108. doi:10.1038/emboj.2012.32

- Simonsen A, Birkeland HC, Gillooly DJ, et al. Alfy, a novel FYVE-domain-containing protein associated with protein granules and autophagic membranes. J Cell Sci. 2004;117(Pt 18):4239-4251. doi:10.1242/jcs.01287
- Suzuki M, Shinohara Y, Ohsaki Y, Fujimoto T. Lipid droplets: size matters. J Electron Microsc (Tokyo). 2011;60 Suppl 1:S101-S116. doi:10.1093/jmicro/dfr016
- Swatek KN, Komander D. Ubiquitin modifications. Cell Res. 2016;26(4):399-422. doi:10.1038/cr.2016.39
- Sztalryd C, Brasaemle DL. The perilipin family of lipid droplet proteins: Gatekeepers of intracellular lipolysis. Biochim Biophys Acta Mol Cell Biol Lipids. 2017;1862(10 Pt B):1221-1232. doi:10.1016/j.bbalip.2017.07.009
- Tanaka K. The proteasome: overview of structure and functions. Proc Jpn Acad Ser B Phys Biol Sci. 2009;85(1):12-36. doi:10.2183/pjab.85.12
- Tansey JT, Sztalryd C, Gruia-Gray J, et al. Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin production, and resistance to diet-induced obesity. Proc Natl Acad Sci U S A. 2001;98(11):6494-6499. doi:10.1073/pnas.101042998
- Tomioka I, Ishibashi H, Minakawa EN, et al. Transgenic Monkey Model of the Polyglutamine Diseases Recapitulating Progressive Neurological Symptoms. eNeuro. 2017;4(2):ENEURO.0250-16.2017. Published 2017 Mar 28. doi:10.1523/ENEURO.0250-16.2017
- Udd B, Partanen J, Halonen P, et al. Tibial muscular dystrophy. Late adult-onset distal myopathy in 66 Finnish patients. Arch Neurol. 1993;50(6):604-608. doi:10.1001/archneur.1993.00540060044015.
- Vadlamudi RK, Joung I, Strominger JL, Shin J. p62, a phosphotyrosine-independent ligand of the SH2 domain of p56lck, belongs to a new class of ubiquitin-binding proteins. J Biol Chem. 1996;271(34):20235-20237. doi:10.1074/jbc.271.34.20235
- van Meer G. Caveolin, cholesterol, and lipid droplets?. J Cell Biol. 2001;152(5):F29-F34. doi:10.1083/jcb.152.5.f29
- Walther TC, Farese RV Jr. The life of lipid droplets. Biochim Biophys Acta. 2009;1791(6):459-466. doi:10.1016/j.bbalip.2008.10.009
- Wang CW. Lipid droplets, lipophagy, and beyond. Biochim Biophys Acta. 2016;1861(8 Pt B):793-805. doi:10.1016/j.bbalip.2015.12.010
- Welander L. Myopathia distalis tarda hereditaria; 249 examined cases in 72 pedigrees. *Acta Med Scand Suppl*. 1951;265:1-124.
- Welte MA. Fat on the move: intracellular motion of lipid droplets. Biochem Soc Trans. 2009;37(Pt 5):991-996. doi:10.1042/BST0370991
- Widrick JJ, Alexander MS, Sanchez B, et al. Muscle dysfunction in a zebrafish model of Duchenne muscular dystrophy. Physiol Genomics. 2016;48(11):850-860. doi:10.1152/physiolgenomics.00088.2016
- Wolins NE, Skinner JR, Schoenfish MJ, Tzekov A, Bensch KG, Bickel PE. Adipocyte protein S3-12 coats nascent lipid droplets. J Biol Chem. 2003;278(39):37713-37721. doi:10.1074/jbc.M304025200
- Wolins NE, Quaynor BK, Skinner JR, Schoenfish MJ, Tzekov A, Bickel PE. S3-12, Adipophilin, and TIP47 package lipid in adipocytes. J Biol Chem. 2005;280(19):19146-19155. doi:10.1074/jbc.M500978200

- Wolins NE, Quaynor BK, Skinner JR, et al. OXPAT/PAT-1 is a PPAR-induced lipid droplet protein that promotes fatty acid utilization. Diabetes. 2006;55(12):3418-3428. doi:10.2337/db06-0399
- Xu G, Sztalryd C, Lu X, et al. Post-translational regulation of adipose differentiation-related protein by the ubiquitin/proteasome pathway. J Biol Chem. 2005;280(52):42841-42847. doi:10.1074/jbc.M506569200
- Yamaguchi T, Matsushita S, Motojima K, Hirose F, Osumi T. MLDP, a novel PAT family protein localized to lipid droplets and enriched in the heart, is regulated by peroxisome proliferator-activated receptor alpha. J Biol Chem. 2006;281(20):14232-14240. doi:10.1074/jbc.M601682200
- Zechner R, Kienesberger PC, Haemmerle G, Zimmermann R, Lass A. Adipose triglyceride lipase and the lipolytic catabolism of cellular fat stores. J Lipid Res. 2009;50(1):3-21. doi:10.1194/jlr.R800031-JLR200
- Zechner R, Madeo F, Kratky D. Cytosolic lipolysis and lipophagy: two sides of the same coin. Nat Rev Mol Cell Biol. 2017;18(11):671-684. doi:10.1038/nrm.2017.76

Acknowledgments

Working at this fascinating yet complex project has been extremely stimulating, allowing me to come in contact with many people whose support and inspiration contribute to my professional and personal growth. The work was carried out initially at the Hospital for Sick Children department of Genetics and Genome Biology in Toronto and then at the IRCCS Fondazione Istituto Nazionale Neurologico C. Besta at the Laboratory of the Muscle Cell Biology in Milan.

First of all I want to thank my past and present supervisors. Dr. Berge Minassian, whose kindness and precious guidance help me to develop a critical thinking, strengthening my self-confidence. I will always be grateful for having had the chance to work in his lab. I also wish to thank Dr. Marina Mora for teaching with profound scientific enthusiasm, all she knew.

My sincere gratitude goes to my PhD tutor Prof. Giuseppe Borsani, our scientific conversations and his kind support have been of paramount importance for the completion of this work.

My unconditional recognition goes to all the people I've met at the bench during these years. They proved to be valued and stimulating colleagues but most important, trustworthy friends. Thanks Peter Peixiang Wang, Xiaochu Zhao, Nivetha Ramachandran, Flavia Blasevich, Cinzia Bragato, Sara Gibertini, Eliana Iannibelli, Simona Saredi and thanks to all the people I've crossed and that I will never forget.

At last, my heartfelt gratitude goes to my family, my husband Carlo, my parents, my brothers and sister and all the friends that with love, supported and encouraged my decisions. I am so proud and grateful for being your wife, daughter, sister and friend. Endless thanks.

Acta Neuropathologica
https://doi.org/10.1007/s00401-020-02164-4

CORRESPONDENCE

Multiomic elucidation of a coding 99-mer repeat-expansion skeletal muscle disease

Alessandra Ruggieri[1,2] · Sergey Naumenko[3] · Martin A. Smith[4,5,6,7] · Eliana Iannibelli[1] · Flavia Blasevich[1] · Cinzia Bragato[1,8] · Sara Gibertini[1] · Kirston Barton[7] · Matthias Vorgerd[9] · Katrin Marcus[10] · Peixiang Wang[11] · Lorenzo Maggi[1] · Renato Mantegazza[1] · James J. Dowling[11] · Rudolf A. Kley[9,12] · Marina Mora[1] · Berge A. Minassian[11,13]

Received: 31 March 2020 / Revised: 15 May 2020 / Accepted: 15 May 2020

Twenty-two individuals across four generations suffer a chromosome 19p13.3-linked autosomal dominant progressive myopathy with distinctive pathology including rimmed ubiquitin-positive autophagic vacuolation [6] (Fig. 1 and Supplementary data). A recombination in the newest (youngest) affected patient (V:13) and repeat linkage analysis on six patients (IV:10, IV:17, III:18, IV:3, IV:23, V:13) refined the disease haplotype to 5.12 Mb containing 164 genes. Sanger (24 genes), whole-exome, whole-genome (Supplementary Table 1) and whole skeletal muscle RNA sequencing proved unrevealing.

Immunohistochemical workup showed that patients' vacuoles and subsarcolemmal regions stained positive for FK2 and p62/SQSTM1 markers, respectively, of ubiquitinated proteins and autophagy (Fig. 1d). The number of stained fibers correlated with clinical severity (Supplementary data). To query the FK2 target(s), we microdissected vacuoles and unaffected myofiber parts for quantitative mass spectrometry. Among the more than 700 identified proteins (Supplementary Table 2), perilipin-4 was the most highly (almost 20-fold) over-represented in vacuoles versus control myofiber regions (Supplementary Fig. 2). Perilipins coat the phospholipid monolayer surrounding lipid droplets and regulate the latter [2]. All five perilipins share an amphipathic domain composed of an 11-amino acid (aa) sequence, which is particularly extensive in perilipin-4, the 11-mer being repeated three times to generate a 33-mer, in turn repeated 29 or 31 times [11] (Supplementary Fig. 3). Perilipin-4 is the most abundantly expressed perilipin in muscle, notwithstanding which, aside from reduced cardiac triacylglycerol

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00401-020-02164-4) contains supplementary material, which is available to authorized users.

✉ Berge A. Minassian
berge.minassian@UTSouthwestern.edu

Alessandra Ruggieri
alessandra.ruggieri@istituto-besta.it

1 Department of Neuroimmunology and Neuromuscular Diseases, Fondazione IRCCS Neurological Institute Carlo Besta, Milan, Italy

2 Department of Molecular and Translation Medicine, Unit of Biology and Genetics, University of Brescia, Brescia, Italy

3 Centre for Computational Medicine, Hospital for Sick Children, Toronto, ON, Canada

4 CHU Sainte-Justine Research Center, Montreal, QC, Canada

5 Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

6 St-Vincent's Clinical School, Faculty of Medicine, UNSW Sydney, Sydney, Australia

7 Garvan Institute for Medical Research, Darlinghurst, NSW, Australia

8 PhD Program in Neuroscience, University of Milano-Bicocca, Monza, Italy

9 Department of Neurology, Heimer Institute for Muscle Research, University Hospital Bergmannsheil, Ruhr-University Bochum, Bochum, Germany

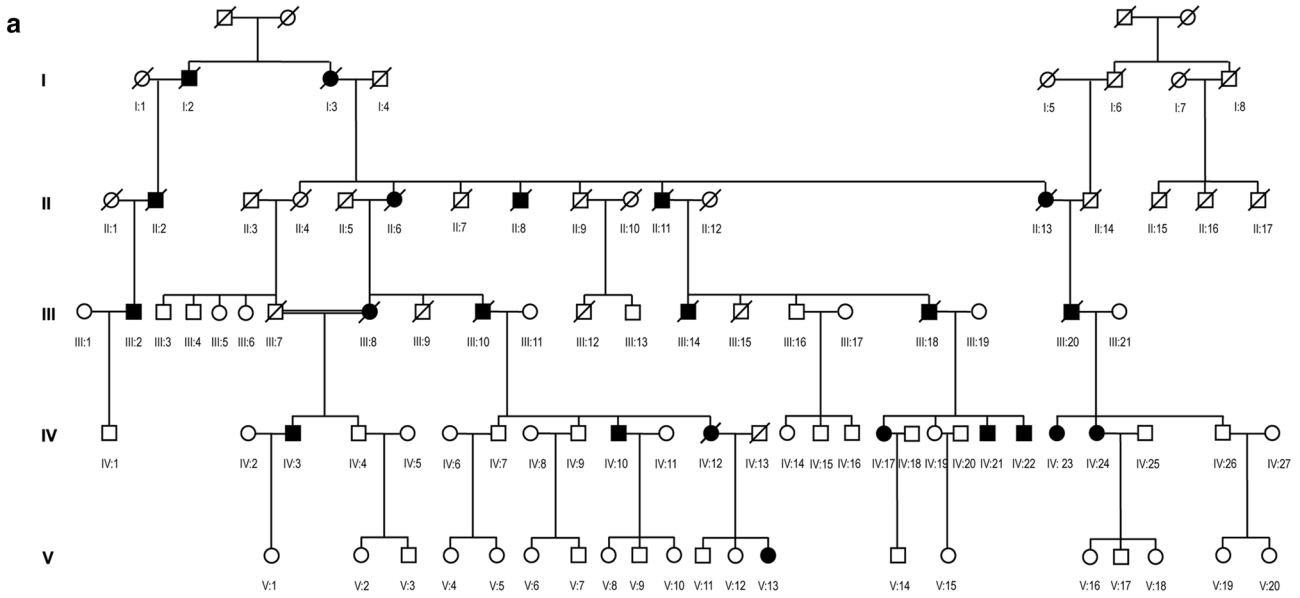
10 Medizinisches Proteom-Center, Ruhr-University Bochum, Bochum, Germany

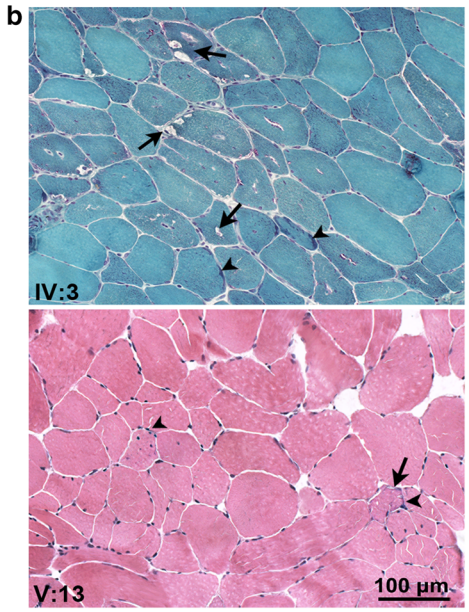
11 Program in Genetics and Genome Biology, Hospital for Sick Children Research Institute, Toronto, ON, Canada

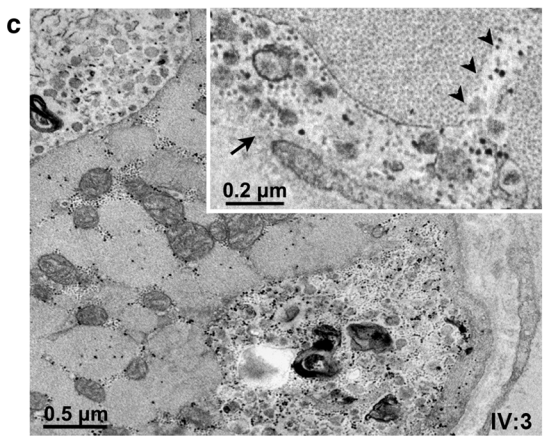
12 Department of Neurology and Clinical Neurophysiology, St. Marien-Hospital Borken, Klinikum Westmuensterland, Borken, Germany

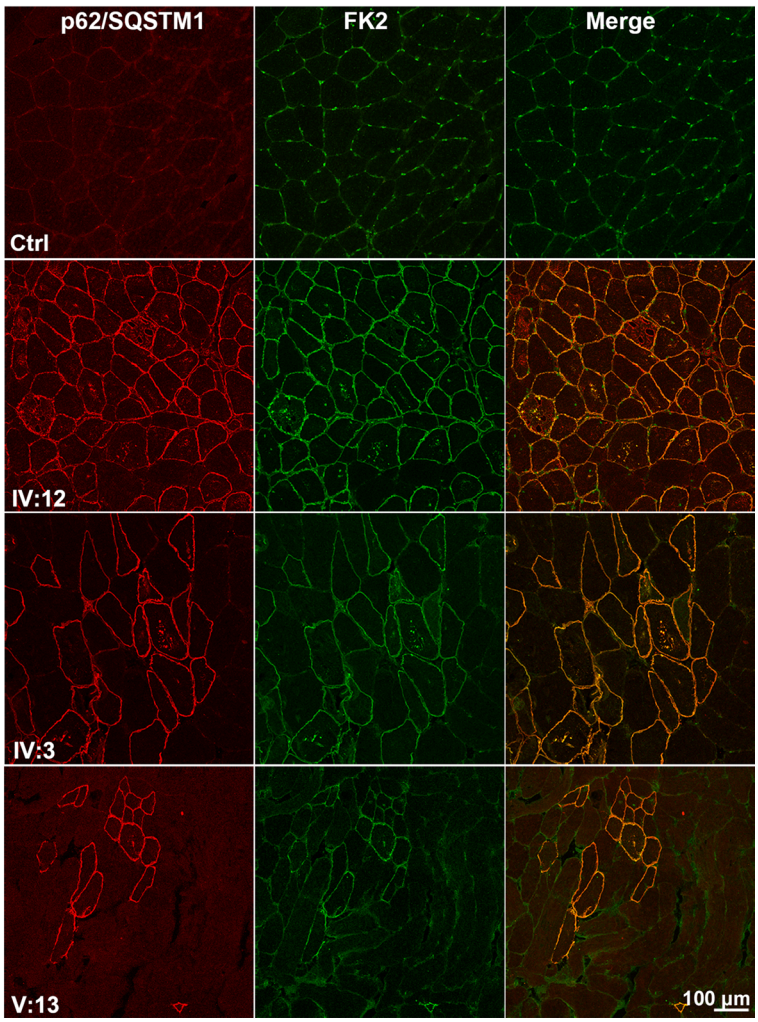
13 Division of Neurology Department of Pediatrics, University of Texas Southwestern, Dallas, TX, USA

Published online: 25 May 2020


a
I
II
III
IV
V
I:1 I:2 I:3 I:4 I:5 I:6 I:7 I:8
II:1 II:2 II:3 II:4 II:5 II:6 II:7 II:8 II:9 II:10 II:11 II:12 II:13 II:14 II:15 II:16 II:17
III:1 III:2 III:3 III:4 III:5 III:6 III:7 III:8 III:9 III:10 III:11 III:12 III:13 III:14 III:15 III:16 III:17 III:18 III:19 III:20 III:21
IV:1 IV:2 IV:3 IV:4 IV:5 IV:6 IV:7 IV:8 IV:9 IV:10 IV:11 IV:12 IV:13 IV:14 IV:15 IV:16 IV:17 IV:18 IV:19 IV:20 IV:21 IV:22 IV: 23 IV:24 IV:25 IV:26 IV:27
V:1 V:2 V:3 V:4 V:5 V:6 V:7 V:8 V:9 V:10 V:11 V:12 V:13 V:14 V:15 V:16 V:17 V:18 V:19 V:20


b
IV:3
V:13
100 µm


c
0.2 µm
0.5 µm
IV:3


d
p62/SQSTM1
FK2
Merge
Ctrl
IV:12
IV:3
V:13
100 µm

◂**Fig. 1** Pedigree and muscle histopathology. **a** Pedigree. **b** Gomori trichrome and H&E staining, respectively, in patients IV:3 and V:3, showing vacuoles (arrows) rimmed, empty, or containing granular or basophilic material (arrowheads) mainly located in the subsarcolemmal region of fibers, fiber size variability, central nuclei and mildly increased endomysial spaces. Note minimal changes in V:13. **c** Electron micrographs unveiling (upper panel) granular debris within a small subsarcolemmal vacuole (arrow) opening to the fiber's surface and sarcolemmal interruption (arrowheads); (lower panel) vacuoles located in the subsarcolemmal region or deep in the sarcoplasm, containing small vesicles, membranous bodies and granular debris. **d** Confocal microscopy of p62/SQSTM1 and FK2 immunostaining showing positivity and almost complete overlap of both proteins in vacuoles and subsarcolemmae of affected fibers, more numerous with increasing clinical severity

levels, its absence in mouse results in no cardiac, skeletal muscle or other impairment [10, 11].

Since *PLIN4* maps to our linked region, we revisited the genomic and transcriptomic patient data and noticed an unusually high coverage in *PLIN4* exon 3 (Supplementary Fig. 4). PCR amplification of this exon in patient genomic DNA and muscle RNA revealed the wild-type band, and a second ~ 1000 bp higher band (Fig. 2a) not present in unaffected relatives or in 60 ethnic controls. The 31 × 33-aa amphipathic domain of perilipin-4 is encoded by 31 × 99 repetitive sequences in exon 3 [4], which poses a computational challenge for aligning short sequencing reads. We amplified cDNA from patient muscle RNA and obtained Oxford Nanopore long-read sequencing, which confirmed that the higher band is an expansion of the normal 31 × 99-nucleotide sequence to 40 × 99 bases, resulting in 297 (9 × 33) extra amino acids (Supplementary Fig. 5). Muscle extract Western blotting with a perilipin-4 antibody showed the presence of a second band consistent in size with the genetic expansion in patients, and absent from controls (Fig. 2b). Immunohistochemistry with the same antibody showed a major increase in perilipin-4 positivity in subsarcolemmal regions and vacuoles of patients compared to controls. The perilipin-4 signal most exactly reproduced the staining with the FK2 (Fig. 2c) and p62/SQSTM1 antibodies (Supplementary Fig. 6). These staining correlated with the diseased muscle fiber type, namely, slow-twitch Type I fibers, known to contain the highest amounts of intramyocellular lipids. Oil Red O staining showed normal lipid content and distribution (Supplementary Fig. 7 and data). Aggrephagy pathway components beyond FK2 and p62/SQSTM1, namely, NBR1 and WDFY3, were upregulated, the former (Fig. 2d) colocalizing with perilipin-4, FK2 and p62/SQSTM1, the latter (Fig. 2e) increased in subsarcolemmae near perilipin-4 positivity but without co-localization. In aggrephagy, p62/SQSTM1 interacts with NBR1, and the two, as an autophagy receptor complex and through their shared LC3-interacting regions, bridge the aggregating ubiquitinated proteins with LC3. Meanwhile, WDFY3 shuttles from the nucleus to the cytoplasm to scaffold the overall structure with PtdIns3P-containing membranes and encapsulate the aggregates in autophagosomes for degradation [8]. The present disease is characterized by dominantly inherited progressively increasing mobilization of aggrephagy at sites of progressive accumulation of a mutated protein, suggesting that the mutation is leading to aggregation, likely through misfolding, exceeding aggrephagic capacity. Continuous formation and fusion of failing aggrephagic vesicles possibly leads to ever larger vacuoles, which disrupt the organization of myofibers and alter their contractile abilities, resulting in atrophy.

Many cases of Inclusion Body Myopathy, the most common of the myopathies, exhibit aggrephagic activation, including NBR1 deposition, not dissimilar to the present patients [5]. The proportion of cases that are due to misfolded proteins, potentially including perilipin-4, remains to be determined.

Perilipin-4 shares its amphipathic domain structure with α-synuclein and exchangeable lipoproteins (ApoA, ApoC and ApoE) [3]. All known mutations (all missense) of α-synuclein and ApoA1 in familial Parkinson disease and amyloidosis, respectively, localize to these proteins' amphipathic regions and transition the repeating helices of these domains to amyloidogenic β pleats [1, 7]. Genomic repeat sequences predispose to expansion [9]. To our knowledge, ours is the first report of an amphipathic domain repeat expansion in disease, and identification of the expansion was only possible with long-read sequencing. The possible occurrence of germline or somatic pathogenic amphipathic region repeat expansions in proteins possessing these domains in their related diseases should be explored.

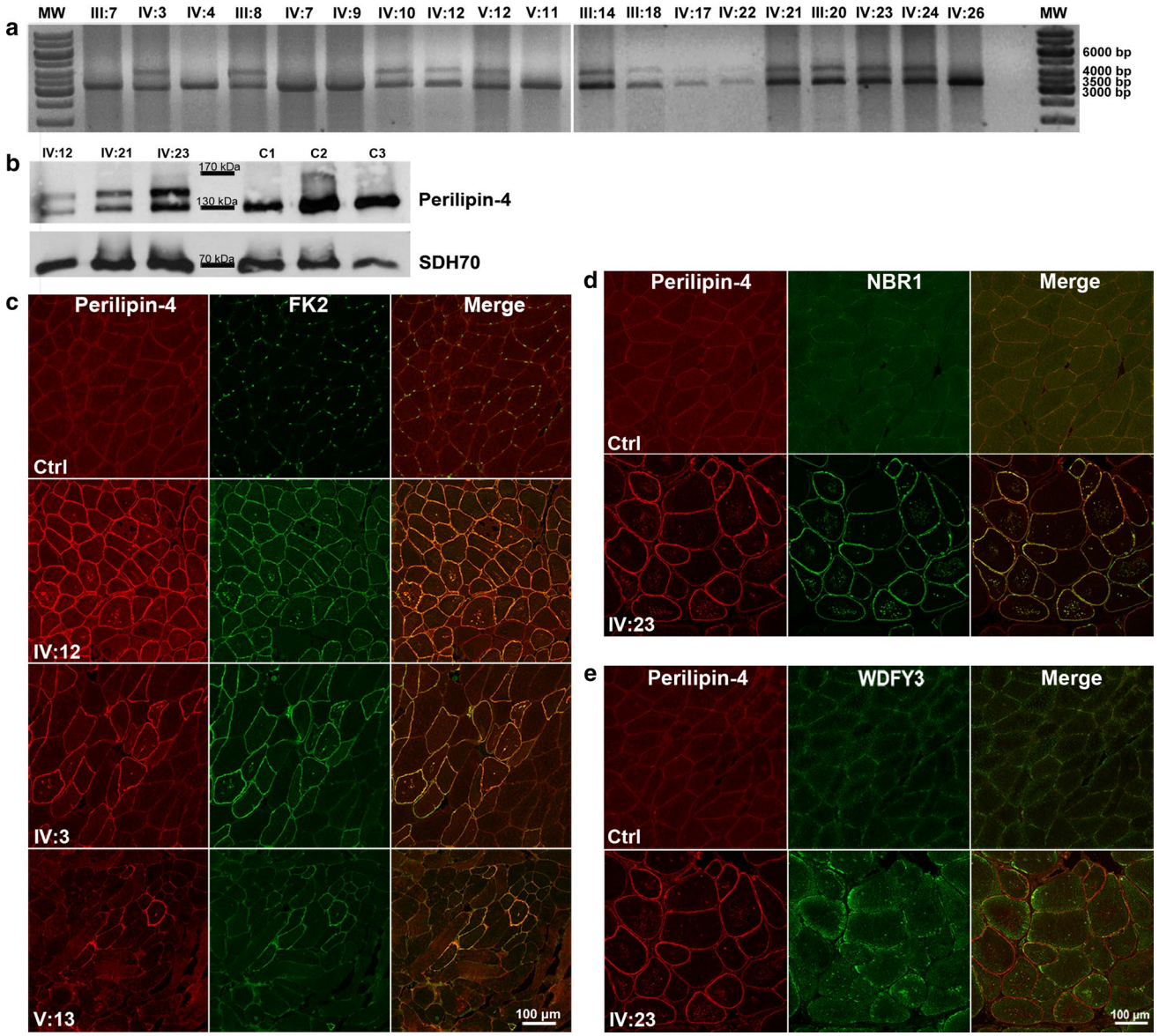


Fig. 2 Perilipin-4 expression and aggrephagy. **a** Exon 3 PCR amplification showing besides the wild-type band, a second one approximately 1000 bp higher only in affected family members. **b** Perilipin-4 western blot revealing a second band in patient muscle, absent in controls. **c** Immunohistochemistry showing perilipin-4 within vacuoles and in the subsarcolemmal region in all affected fibers, overlapping with FK2 by confocal microscopy analysis. **d**, **e** Co-localization of perilipin-4 and aggrephagy-related proteins NBR1 (**d**) and WDFY3 (**e**), showing upregulation of both in patient muscle, with good overlap of NBR1-perilipin-4and increase of WDFY3 near perilipin-4

Acknowledgements We would like to thank the patients and the extended family for constant support and cooperation. We also gratefully acknowledge EuroBioBank and the Telethon Network of Genetic Biobanks for providing biological samples; Dr. Lucia Morandi for her past collaboration with acquisition of early clinical data and Dr. Simona Saredi for her past support with experiments. The family wishes to dedicate this work to their sister and mother Anna, affected by this disease and recently deceased: "she taught us the importance of perseverance and the strength of optimism".

Author contributions AR, MM and BAM conceived and designed the study and wrote the manuscript with input from all authors. MM and BAM supervised and provided critical discussion of the data. AR, SN, MAS and KB performed genetic experiments and/or interpreted data. MV, KM and RAK preformed proteomic experiments and/or interpreted data. AR, EI, FB, BC and GS performed molecular, immunohistochemistry, immunoblot, confocal and electron microscopy experiments. LM and RM acquired and analyzed clinical data. PW and JJD provided experimental support and advice.

References

1. Arciello A, Piccoli R, Monti DM (2016) Apolipoprotein A-I: the dual face of a protein. FEBS Lett 590:4171–4179. https://doi.org/10.1002/1873-3468.12468
2. Bosma M (2016) Lipid droplet dynamics in skeletal muscle. Exp Cell Res 340:180–186. https://doi.org/10.1016/j.yexcr.2015.10.023
3. Bussell R, Eliezer D (2003) A structural and functional role for 11-mer repeats in alpha-synuclein and other exchangeable lipid binding proteins. J Mol Biol 329:763–778. https://doi.org/10.1016/s0022-2836(03)00520-5
4. Čopič A, Antoine-Bally S, Giménez-Andrés M, La Torre GC, Antonny B, Manni MM et al (2018) A giant amphipathic helix from a perilipin that is adapted for coating lipid droplets. Nat Commun 9:1332. https://doi.org/10.1038/s41467-018-03717-8
5. D'Agostino C, Nogalska A, Cacciottolo M, Engel WK, Askanas V (2011) Abnormalities of NBR1, a novel autophagy-associated protein, in muscle fibers of sporadic inclusion-body myositis. Acta Neuropathol 122:627–636. https://doi.org/10.1007/s00401-011-0874-3
6. Di Blasi C, Moghadaszadeh B, Ciano C, Negri T, Giavazzi A, Cornelio F et al (2004) Abnormal lysosomal and ubiquitin-proteasome pathways in 19p13.3 distal myopathy. Ann Neurol 56:133–138. https://doi.org/10.1002/ana.20158
7. Flagmeier P, Meisl G, Vendruscolo M, Knowles TP, Dobson CM, Buell AK et al (2016) Mutations associated with familial Parkinson's disease alter the initiation and amplification steps of α-synuclein aggregation. Proc Natl Acad Sci USA 113:10328–10333. https://doi.org/10.1073/pnas.1604645113
8. Knaevelsrud H, Simonsen A (2010) Fighting disease by selective autophagy of aggregate-prone proteins. FEBS Lett 584:2635–2645. https://doi.org/10.1016/j.febslet.2010.04.041
9. Mirkin SM (2006) DNA structures, repeat expansions and human hereditary disorders. Curr Opin Struct Biol 16:351–358. https://doi.org/10.1016/j.sbi.2006.05.004
10. Pourteymour S, Lee S, Langleite TM, Eckardt K, Hjorth M, Bindesbøll C et al (2015) Perilipin 4 in human skeletal muscle: localization and effect of physical activity. Physiol Rep 3:e12481. https://doi.org/10.14814/phy2.12481
11. Sztalryd C, Brasaemle DL (2017) The perilipin family of lipid droplet proteins: gatekeepers of intracellular lipolysis. Biochim Biophys Acta Mol Cell Biol Lipids 1862:1221–1232. https://doi.org/10.1016/j.bbalip.2017.07.009

SUPPLEMENTARY MATERIAL Acta Neuropathologica

Multiomic elucidation of a coding 99-mer repeat-expansion skeletal muscle disease

Alessandra Ruggieri, Sergey Naumenko, Martin A. Smith, Eliana Iannibelli, Flavia Blasevich, Cinzia Bragato, Sara Gibertini, Kirston Barton, Matthias Vorgerd, Katrin Marcus, Peixiang Wang, Lorenzo Maggi, Renato Mantegazza, James J. Dowling, Rudolf A. Kley, Marina Mora, Berge A. Minassian

Correspondence:

Berge A. Minassian
berge.minassian@UTSouthwestern.edu

Alessandra Ruggieri
alessandra.ruggieri@istituto-besta.it

SUPPLEMENTARY DATA

Morphological findings

Pathological features of muscle were: markedly variable fiber size with hypertrophic and angulated atrophic fibers, mild to moderate increase in endomysial and perimysial fat and fibrous connective tissue, central nuclei, fiber splitting, and vacuoles. Vacuoles were often rimmed and appeared either empty or filled with basophilic and/or granular material. They were present either at the inner of the fibers where their size was variable, sometimes replacing most of the sarcoplasm; or, more often, they were clustered, multiple and small in size, at the fiber surface (Fig. 1b and Supplementary Fig. 1a).

The severity of the histopathologic changes varied between patients and correlated with disease severity (Fig.1b and Supplementary Fig.1a and 2a). In particular a muscle biopsy in the youngest affected case (V:13) showed minimal changes such as variation in fiber size, a few central nuclei, and basophilic underlining of the surface in rare fibers (Fig. 1b).

The electron microscopy analysis revealed that vacuoles were often membrane-limited, and, especially those at the fiber surface, were also bordered by the basal lamina. Vacuoles contained numerous membranous bodies, partially degraded organelles, and amorphous or granular material (Supplementary Fig. 1b,c,d,f). They were located either within the cytoplasm where they were usually large in size, and at the fiber surface, where they were smaller in size and either single or multiple, often abutting the extracellular space (Supplementary Fig. 1b,c,f) (see also previous Di Blasi et al.'s report [3]). These peripheral multiple microvacuoles likely correspond to the subsarcolemmal regions positive for Perilipin-4, ubiquitinated proteins and autophagic markers seen in light microscopy. The basal lamina, in addition to limiting many vacuole surfaces, often formed multiple layers and loops in the extracellular space facing the vacuoles (Supplementary Fig.1b,c). Glycogen was abundant in the intermyofibrillar spaces and close to and within vacuoles. Tubulo-filamentous inclusions were not observed in the sarcoplasm or in nuclei, nor myofibrillar, or mitochondrial alterations. Sarcoplasmic reticulum cisternae appeared sometimes enlarged especially at the triads (Supplementary Fig. 1b,f); lipid droplets were rare, appearing normally empty or, in a few cases, with a dense core (Supplementary Fig. 1b).

Percentages of fibers with increased positivity to perilipin-4 seemed to be well correlating with the patients' phenotype severity measured according to the Walton and Gardner-Medwin scale, as shown in the Supplementary Fig. 2a graph which only included those patients for whom neurological examination was performed at time of biopsy.

The number of perilipin-4- or FK2-positve fibers directly correlated with clinical severity, ranging from very few in patients IV:17 and V:13, both practically asymptomatic, to all fibers in the most severely affected patients (Supplementary Fig.2).

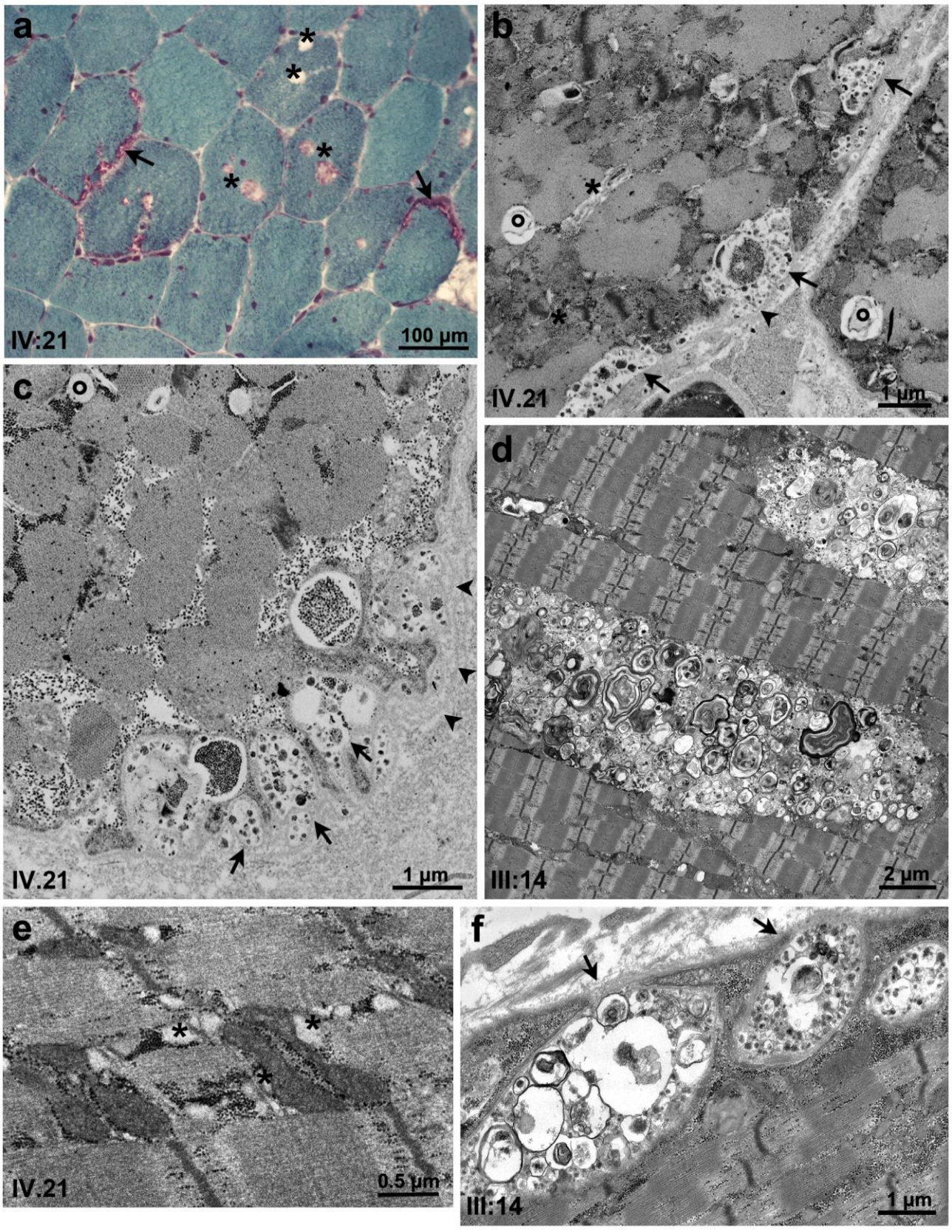
Perilipin-4 immunostaining in muscle biopsies from genetically determined vacuolar myopathies such as DNAJB6-, GNE-, VMA21-, LAMP2-, or GAA-mutated myopathies showed lack of positivity in the vacuoles and in the subsarcolemmal region of the fibers, except in a few spots due to autofluorescence (Supplementary Fig. 2),

Quantitation of co-localization between perilipin-4 and NBR1, performed in patient IV:3, revealed a significant increase of co-localization signal compared to a control (a VCP-mutated patient with autophagic vacuoles) (IV:3 = 0.7427 ± 0.02107, vs Ctrl = 0.3130 ± 0.02170, p=0.0001). Similarly, perilipin-4 and FK2 co-localization signal (IV:3 = 0.7473 ± 0.01617, vs Ctrl: = 0.2803 ± 0.001333, $p<0.0001$), and perilipin-4 and p62 co-localization signal (IV:3 = 0.7070 ± 0.00493, vs Ctrl = 0.2923 ± 0.03308, p=0,0002) were significantly increased in the patient.

SUPPLEMENTARY FIGURES

Supplementary Fig. 1 Muscle morphological features

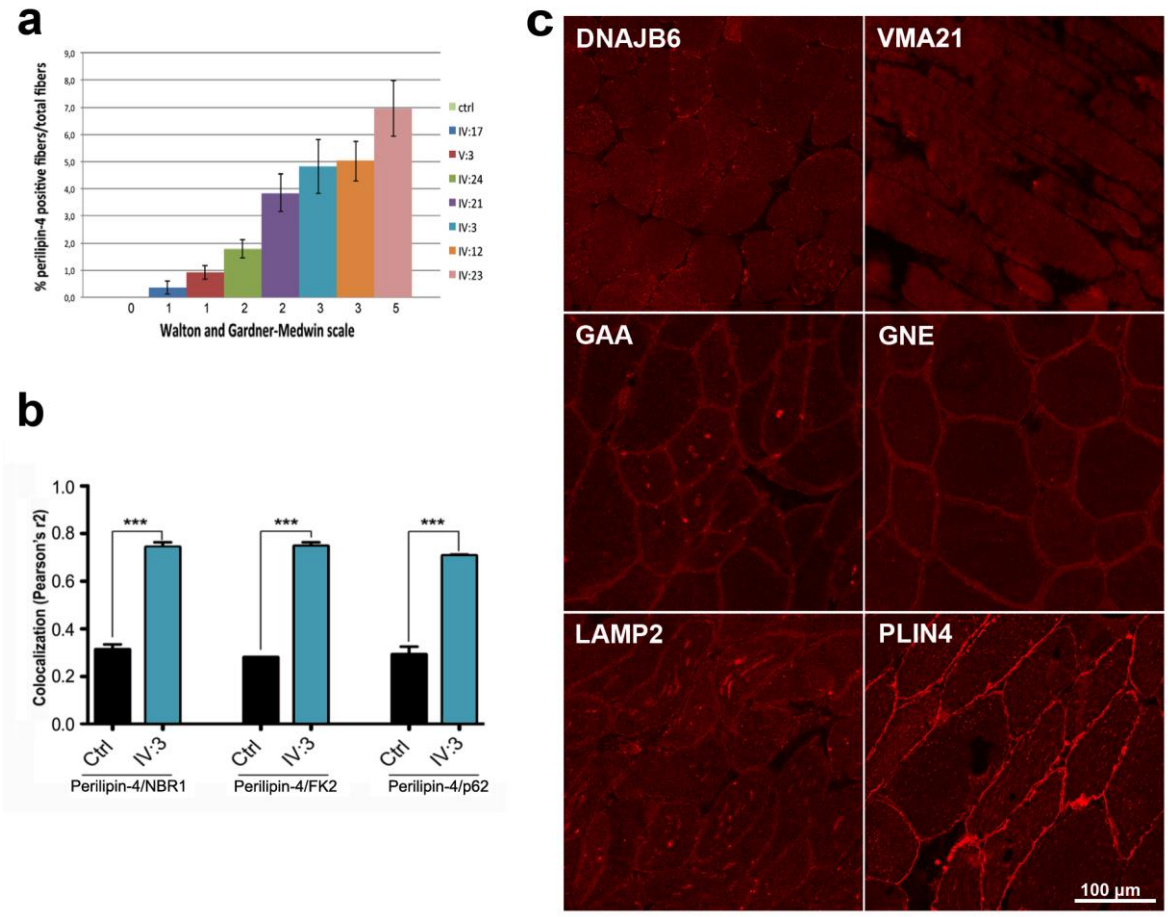
(a) Gomori trichrome staining of patient IV:21 muscle showing presence of numerous vacuoles at the fiber surfaces which appear rimmed and filled with granular basophilic material (arrows),



and centrally located vacuoles empty or containing slightly basophilic material (asterisks). (b-f) Electron microscopy images of transverse and longitudinal ultrathin muscle sections showing: cytoplasmic and peripheral vacuoles sometimes membrane-limited, filled with membranous bodies, partially degraded organelles, and amorphous or granular material (b,c,d,f); peripheral multiple vacuoles which are small in size and often abutting the extracellular space (b,c). Some of these vacuoles are partially or totally delimited only by the basal lamina (arrows) that also forms multiple layers and proliferating loops in the extracellular space facing the vacuoles (arrowheads). (b,c,d,f) Glycogen particles are present in the intermyofibrillar spaces and within vacuoles; (b,f) sarcoplasmic reticulum cisternae and triads appear enlarged (asterisks); (b,c) lipid droplets are rare, and sometimes with a dense core (circles).

Supplementary Fig 2 Quantitation of perlipin-4-positive fibers and of its co-localization with NBR1, FK2, p62, and perilipin-4 immunostaining of in genetically determined vacuolar myopathies

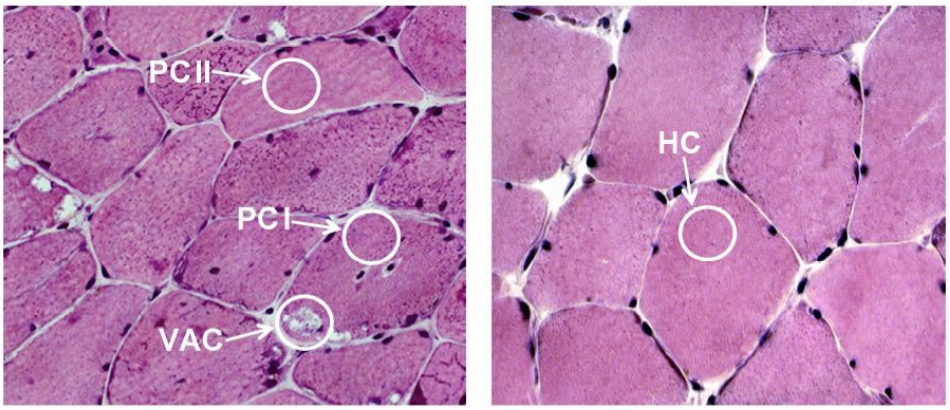
(a) Graph showing quantitation of perilipin-4-positive fibers relative to clinical severity, evaluated according to the Walton and Gardner-Medwin scale, at the time of muscle biopsy. For each patient the scale's values are reported. (b) quantitation of perilipin-4 signal co-localization with NBR1, FK2 and p62, showing significant increased colocalization in patient IV:3 compared to a control. Comparisons between the two groups were done by two-tailed unpaired t-test. *** indicates $p \leq 0.0002$ (c) Perilipn-4 immunostaining in muscle from genetically determined vacuolar myopathies showing lack of positivity in vacuoles and normal low positivity in subsarcolemmal regions of the fibers. The few spots apparently postitive in LAMP2- and GAA-mutated patient muscles are autofluorescent lysosomal corpuscles.



Supplementary Fig 3 Laser microdissection proteomic analysis

(a) Representative images showing how equivalent portions of muscle fibers, either containing rimmed vacuoles (labeled VAC), or the sarcoplasm of the same vacuolated fiber (PCI, patient control I), or a portion of a healthy-looking fiber (PCII), or of a healthy control muscle fiber (HC) were collected by laser micro-dissection for proteomic analysis. (b) Relative ratios of perilipin-4 content in the various portions of the fibers, after proteomic analysis.

a

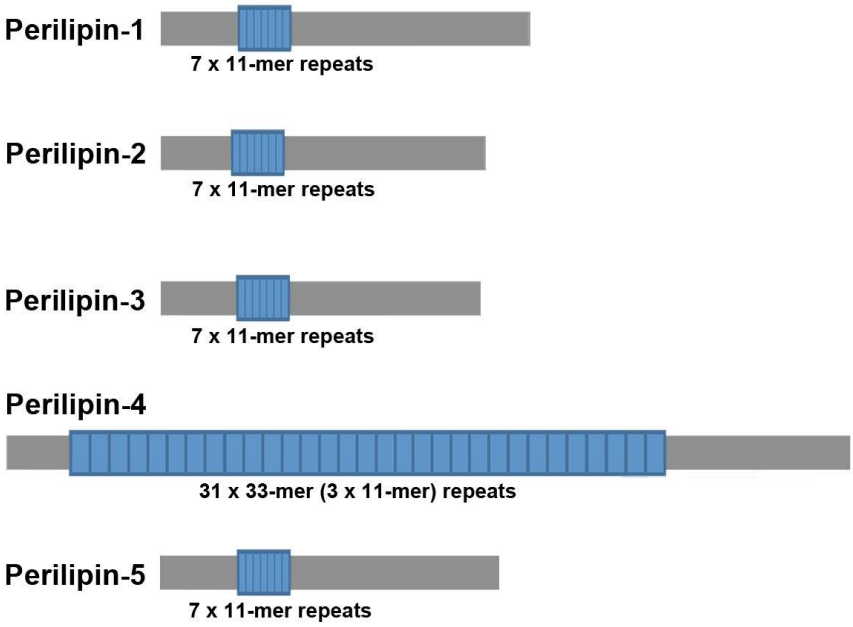


b

Accession	Locus	Ratio VAC / HC
Q96Q06 Perilipin-4	chr19:4502192-4517716	19,57080292
Accession	**Locus**	**Ratio VAC / PC II**
Q96Q06 Perilipin-4	chr19:4502192-4517716	19,57080292
Accession	**Locus**	**Ratio VAC / PC I**
Q96Q06 Perilipin-4	chr19:4502192-4517716	0,850904475
Accession	**Locus**	**Ratio PC I / PC II**
Q96Q06 Perilipin-4	chr19:4502192-4517716	23

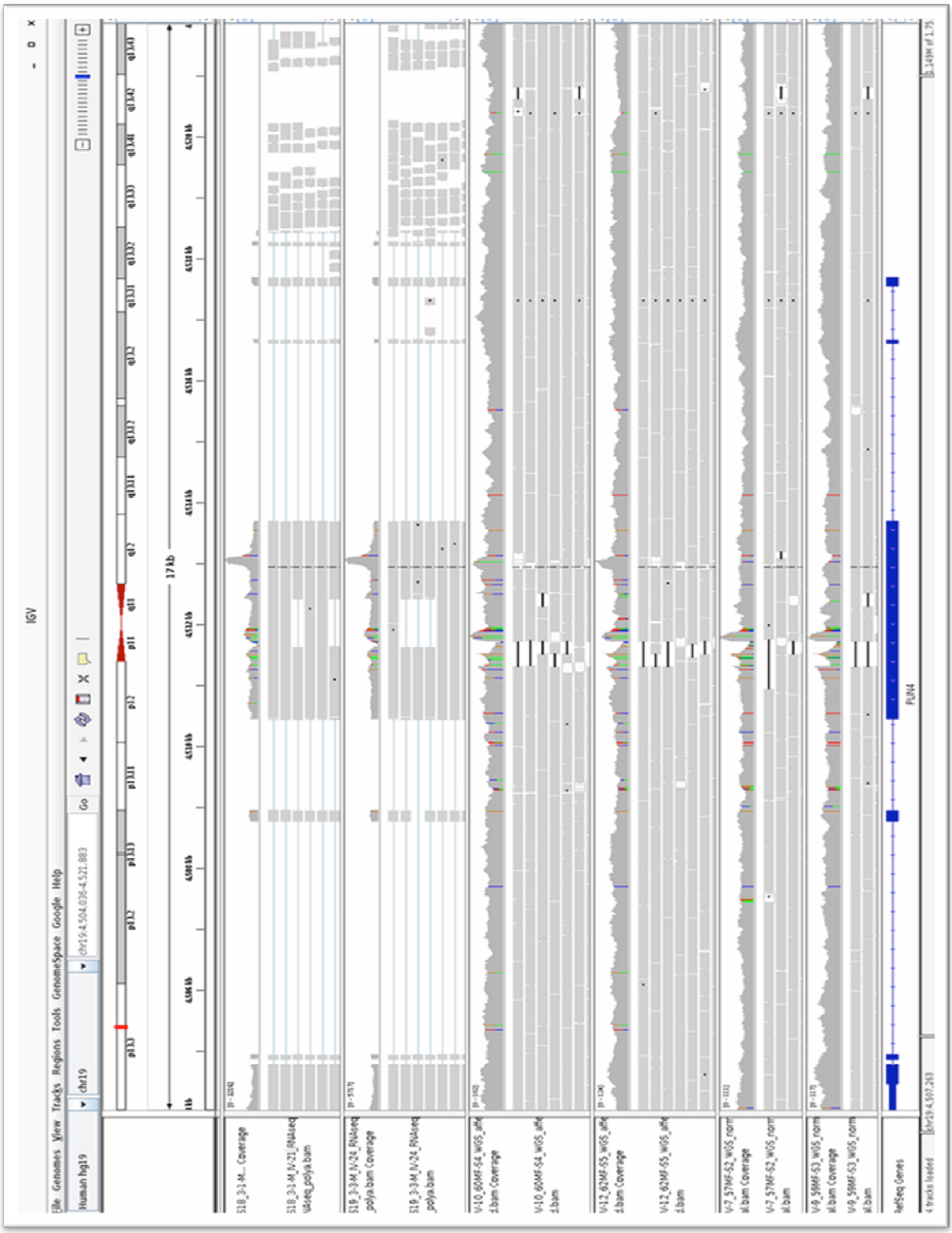
Supplementary Fig. 4 Schematic representation of perilipins

As mentioned in the text, the perilipin-4 33-mer repeat sequence has a normal variation in the human population of 29 or 31 repeats (https://www.uniprot.org/uniprot/Q96Q06)



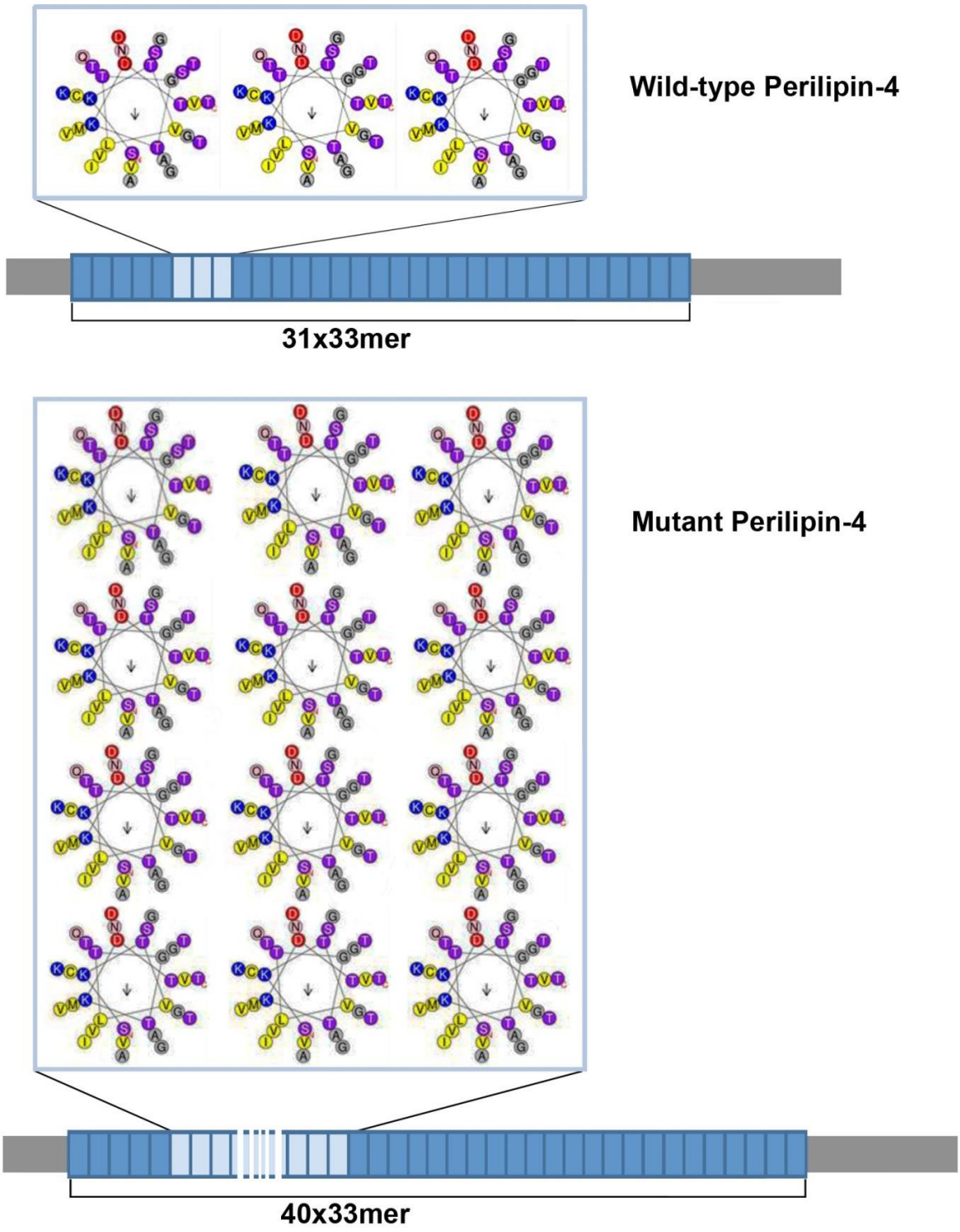
Supplementary Fig. 5 Next generation sequencing

IGV (Integrative Genomics Viewer) [9] visualization of next generation sequencing results, both RNA (rows 1, 2) and genome (rows 3-6) views showing presence in patients of a high peak of coverage (rows 1-4) in PLIN4 exon 3 not present in the healthy siblings (row 5, 6).



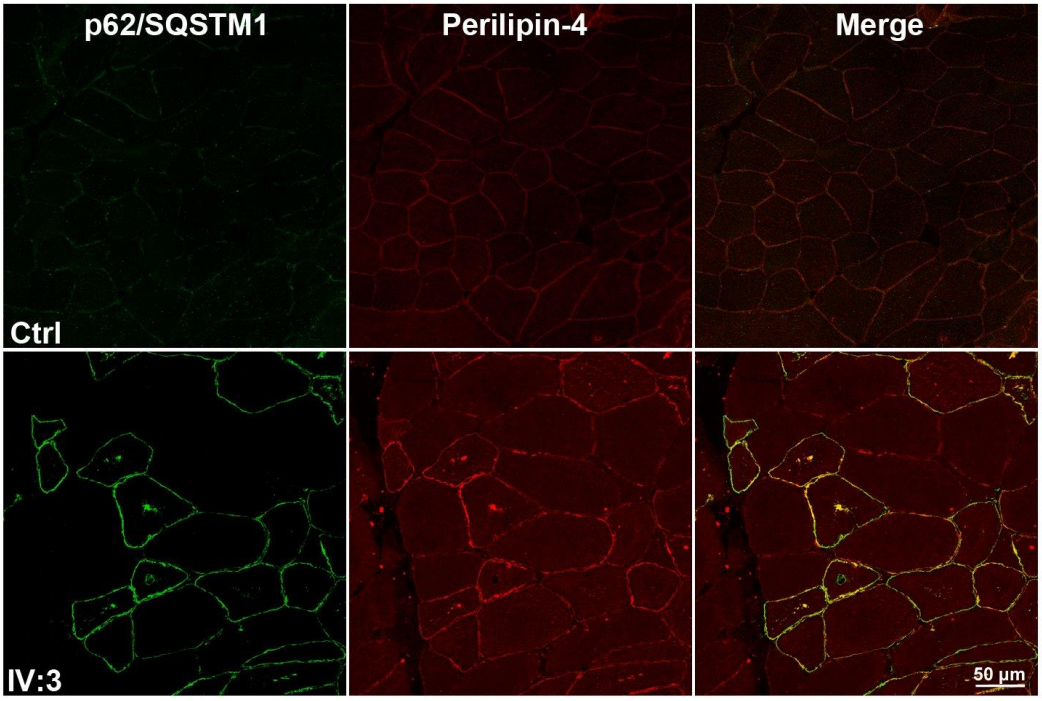
Supplementary Fig. 6 Perilipin-4 structure

Schematic representation of the 33-mer wild-type amphipathic region and its expansion in the mutant protein. The 33-mer helices were sketched using Heliquest [4].



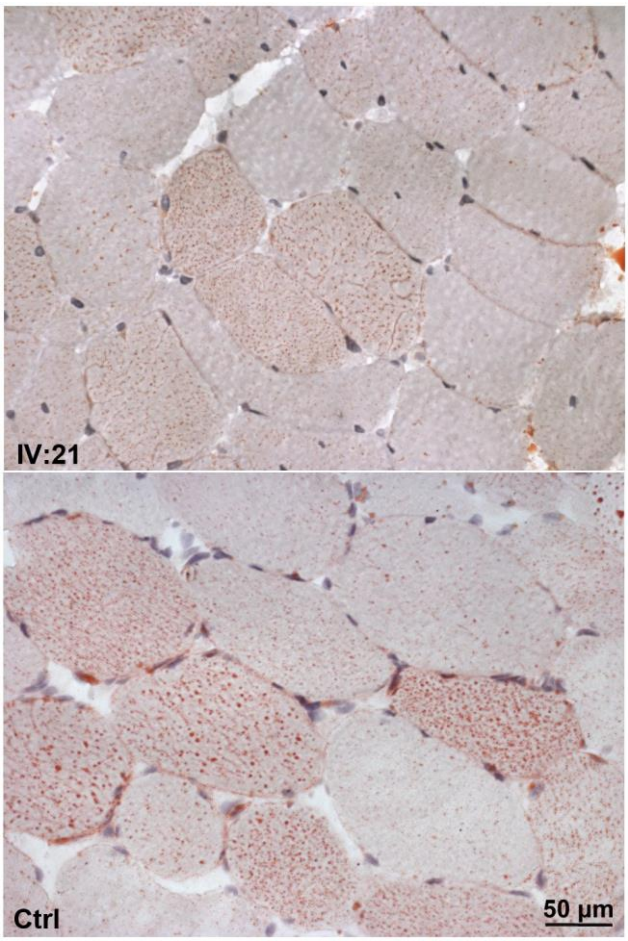
Supplementary Fig. 7 Perlipin-4 and p62/SQSTM1 co-localization

Immunohistochemistry showing upregulation of p62/SQSTM1 and its co-localization with perilipin-4 in vacuoles and fiber subsarcolemmal regions in a patient, but not in control muscle.



Supplementary Fig. 8. Oil Red O staining.

Oil Red O staining, showing similar lipid content in a patient muscle biopsy compared to a control.



MATERIALS AND METHODS

Patients:

Patients were studied with informed consent to use biological samples for diagnosis and research. Investigations were conducted in accordance with protocols approved by the institutional review boards of the C. Besta Neurological Institute of Milan and The Hospital for Sick Children of Toronto.

In all patients creatine kinase was normal or slightly increased. Electromyography showed myopathic potentials in proximal and distal limb muscles in all affected individuals and frequent repetitive high frequency discharges. Clinical features, initially affecting distal lower limb muscles, progressed over time with weakness of scapular muscles.

Examination of the asymptomatic patient (V:13) at the age of 32 years revealed, on the Medical Research Council (MRC) scale, hand extensor (4/5), neck flexor (4/5) and hallux extensor (4/5) muscle weakness. Electromyography showed positive repetitive complex discharges at needle insertion and myopathic pattern, particularly in tibialis anterior and biceps brachialis muscles. Biopsy of quadriceps muscle showed some centralized nuclei and vacuoles in three myofibers.

Morphological studies:

Biopsies of the quadriceps muscles from patients IV:3, IV:10, IV:21, IV:23, IV:24, and V:13 were recently obtained, snap-frozen in isopentane/liquid nitrogen, and maintained in liquid nitrogen. These muscle samples, together with those already available from patients, IV:12 and IV:17, were processed for morphological studies. Transverse cryostat sections were routinely stained for: haematoxylin-eosin, Gomori trichrome, periodic acid-Schiff, Oil Red O, myofibrillar ATPase, acid phosphatase, NADH dehydrogenase, cytochrome C oxidase, succinic dehydrogenase. Immunofluorescence was performed using primary antibodies to: poly-ubiquitinylated proteins (monoclonal PW8810, clone FK2 from Biomol, Enzo Life Sciences, Inc. Farmingdale, USA; 1:100), p62 (guinea pig polyclonal GP62-c, Progene Biotechnik, Heidelberg, Germany; 1:100), p62 (rabbit polyclonal P0067, Merck Sigma-Aldrich, Darmstadt, Germany 1:100) Perilipin-4 (guinea pig polyclonal GP34, Progene Biotechnik, Heidelberg, Germany; 1:100), NBR1 (mouse monoclonal 4BR, Santa Cruz Biotechnology Inc, Santa Cruz CA, USA; 1:100), WDFY3 (mouse monoclonal B-4, Santa Cruz Biotechnology Inc, Santa Cruz CA, USA; 1:100). Cryosections were incubated in the primary antibodies for 2hr, followed by incubation in the secondary antibodies anti-guinea pig, or anti-mouse IgG,Alexa 555 or Alexa 488 conjugated (Invitrogen Life

Technologies, Carlsbad, CA, USA; 1:1500), for 1 hr. Muscle sections were examined under either a Zeiss Axioplan fluorescence microscope (Carl Zeiss AG, Oberkochen, Germany) or a Leica confocal microscope equipped with hybrid and argon lasers (Leica Microsystems, Wetzlar, Germany). Small fragments of muscle tissues, fixed in 4% glutaraldehyde and embedded in Epon 812 resin (Electron Microscopy Sciences, Hatfield PA, USA), were stained with UranyLess (Electron Microscopy Sciences, Hatfield PA, USA) and examined under a FEI Technai electron microscope.

To quantify percentages of fibers positive to perilipin-4 counts were done on pictures of muscle biopsies (at least 3 fields), taken at 20X magnification. To quantify colocalization of perilipin-4/NBR1, perilipin-4/FK2, and perilipin-4/p62, the Fiji plugin JaCoP (JustanotherColocalizationPlugin) analysis was used [2]. The measurement of perilipin-4 overlapping with NBR1, FK2 and p62 was calculated on the entire field of three different images of muscle tissue from a patient and a control (a VCP-mutated patient with autophagic vacuoles). Thresholds were automatically calculated by the software and not set by the operator to avoid biased data. The co-localization measures were reported as Pearson's r^2 units, and comparisons between the two groups were done by two-tailed unpaired t-test.

Western Blot analysis:

Muscle tissue fragments were homogenized in a buffer containing 20 mM Tris (pH 7.8), 137 mM NaCl, 2.7 mM KCl, 1 mM MgCl2, 1% Triton X-100, 10% (w/v) glycerol, 1 mM EDTA, 1 mM dithiothreitol supplemented with protease inhibitor cocktail (P8340, Merck Sigma-Aldrich, Darmstadt, Germany) and incubated 1h at 4°C. The samples were then centrifuged at 13,000 rpm for 10 min and the supernatants collected. Protein content was measured with the DC protein assay (Bio-Rad). Samples run on a 7.5% or 4-15% acrylamide gels, were transferred into a nitrocellulose membrane for immunological detection using the following primary antibodies: perilipin-4 (guinea pig polyclonal AP33136SU-N, OriGene, Rockville, USA) and SDH70 (anti-complex II 70 kDa Fp subunit monoclonal antibody, cod 459200, Invitrogen, Thermo Fisher Scientific, Carlsbad, California, USA) as loading control.

Linkage analysis:

A combination of Illumina chip arrays was used to generate data for whole genome linkage analysis. In detail, SNPs from Illumina Omni1 chip formerly used for CNV analysis on two patients (IV:10 and IV:17) were combined with SNPs from Illumina Omni2.5 array run for patients III:18, IV:3, IV:23 and V:13. Linkage dysequilibrium-based SNP pruning on communal

SNPs was then performed using PLINK software [8] and MERLIN software (version 1.1.2) [1] was used for parametric linkage analysis with parameters considering a rare, dominant and fully penetrant disorder (0.0001, 1.0, 1.0). To further define the exact boundaries of the haplotype shared among the affected members of the family, we combined our previous microsatellite analysis results with the SNPs array data.

Genome sequencing and variants filtering:

Library preparation and subsequent sequencing were performed at The Centre for Applied Genomics in Toronto, following Illumina TruSeq Nano DNA library preparation protocol. In brief, indexed TruSeq Illumina adapters with overhang-T were added to the A-tailed fragments of DNA obtained with a Covaris LE220 instrument. Six cycles of PCR were performed for the enrichment of the libraries and were validated with Bioanalyzer DNA High Sensitivity chip and quantified by qPCR using Kapa Library Quantification Illumina/ABI Prism Kit protocol (KAPA Biosystems). The sequencing of the equimolar pooled libraries was performed on an Illumina HiSeq X platform following Illumina's protocol, generating paired-end reads of 150-bases in length. The variants obtained were filtered based on the linkage analysis results, keeping only the heterozygous ones present in both affected patients and not shared by the unaffected siblings. Moreover, the remaining variants were filtered for frequency, excluding those with a minor allele frequency >0.01 based on 1000genomes, ExAc, ESP and ClinVar databases. Genome coordinates refer to reference genome assembly GRCh37 hg19.

RNAseq:

For the analysis, two samples were prepared with the NEBNext Ultra Directional RNA library prep kit for Illumina with poly(A) enrichment. After verifying RNA quality and quantity, 500ng of total RNA were enriched for poly-A mRNA, retrotranscribed to double stranded cDNA and modified for Illumina adapters ligation. Amplification of the library fragments was executed adding a different barcode adapter at each step, to allow for multiplex sequencing. Libraries pooled in equimolar quantities were pair-end sequenced on an Illumina HiSeq 2500 platform following Illumina's protocol to generate paired-end reads of 126-bases in length.

Laser microdissection, protein digestion and mass spectrometry:

Frozen sections from two affected samples (IV:12 and IV:21) were collected on polyethylene terephthalate membranes (Leica Microsystems) and stained with hematoxylin and eosin (H&E). By laser microdissection using LMD 6500 (Leica Microsystems), different tissue's areas (a total

of 250,000 mm2 areas each) were collected into tubes containing 40 µl of formic acid. The areas were selected as following: rimmed vacuoles (defined as VAC), an equivalent portion of the sarcolemma of the vacuolated fibers (defined as PCI, patient control I), and an equivalent portion of the sarcolemma of healthy-looking fibers (defined as PCII, patient control II) (See Supplementary Fig. 1a). A tryptic in solution-digestion was then carried out as described, followed by nano high-performance liquid chromatography-mass spectrometry (nanoHPLC-ESI-MS/MS) [5].

Relative protein quantification:

Using the Mascot search engine (Matrixscience, London, UK) with parameters as previously reported [7], the data obtained by ESI-MS/MS analysis mass spectrometry were searched against the entire Uniprot/Swissprot (release 2014/10, 546,790 entries) database. Evaluation of the quality of identified peptide spectrum matches and protein quantification were performed as reported [7], and the obtained data were normalized by expressing in percentage the proportion of spectral counts for each protein versus the total sum of spectral counts in the individual samples. For each protein, the ratios between the averaged proportions in rimmed vacuoles and control samples were calculated and statistical significance was evaluated with a two-tailed unpaired t-test (equal variances assumed).

Exon 3 length evaluation by PCR amplification:

100ng of genomic DNA from all the available family members, were amplified using the following primers 3F-ACCACTCGGTCTGCACTTAand 3R-CCTCCAACTCATTCTGCAGC. In detail, a 20µl of the reaction mixture included: 1x Buffer Phusion GC (F519L, Thermo Fisher Scientific, Carlsbad, California, USA), 200µM each dNTPs (dNTPs mix, 11581295001, Roche Merck, Darmstadt, Germany), 0.5µM each forward and reverse primers, 3% DMSO, 0.4µl Phire Hot Start II DNA Polymerase (F122S, Thermo Fisher Scientific, Carlsbad, CA, USA). The amplified products were subsequently run on a 1% agarose gel and the presence of a single (wt) or two (wt and mut) bands was observed under a UV transilluminator equipped with a Kodak EDAS 290 camera (Eastman Kodak Company, Rochester NY, USA).

Oxford Nanopore long read sequencing:

Nanopore sequencing was performed on 500ng of linearized plasmids (pCMV6-AN-GFP) containing either the wild type or the mutant PLIN4 allele of patient cDNA retrotranscribed from

total muscle RNA extract. Both samples were multiplexed with the Native Barcoding kit (EXP-NBD103, Oxford Nanopore Technologies) then prepared for sequencing with the 1D Ligation Sequencing Kit (SQK-LSK108, Oxford Nanopore Technologies) using the manufacturer's recommendations. Base calling and demultiplexing were performed offline with Guppy version 3.2.4 (Oxford Nanopore), using optional parameters '--chunks_per_runner 1500' in high accuracy mode for base calling and '--trim_barcodes --num_extra_bases_trim 30 --barcode_kits "EXP-NBD103"' for demultiplexing. Demultiplexed reads were then filtered for full-length plasmids within an expected size range of 10-12.5 kilobases, resulting in 1121 (wt) and 1164 (mut) full-length reads, which were then subjected to de novo assembly using Canu version 1.8 [5] using options 'genomeSize=12k corMhapOptions="--threshold 0.8 --ordered-sketch-size 1000 --ordered-kmer-size 14" corrected Error Rate=0.105'. The contig assembled from the most reads was then aligned to the human reference genome (GRCh38/hg38) using minimap2 version 2.17 [6] with options '-x splice' to extract the inserted cDNA sequence. The extracted sequence was then polished using Medaka version 10.1 (Oxford Nanopore Technologies) using parameters '-m r941_min_high'.

References

1. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97-101. https://doi:10.1038/ng786.
2. Bolte S, Cordelières FP (2006) A guided tour into subcellular colocalization analysis in light microscopy. J Microsc. 224(Pt 3):213-32. https://doi:10.1111/j.1365-2818.2006.01706.x.
3. Di Blasi C, Moghadaszadeh B, Ciano C, Negri T, Giavazzi A, Cornelio F et al. (2004) Abnormal lysosomal and ubiquitin-proteasome pathways in 19p13.3 distal myopathy. Ann Neurol 56:133-8. https://doi:10.1002/ana.20158.
4. Gautier R, Douguet D, Antonny B, Drin G (2008) HELIQUEST: a web server to screen sequences with specific α-helical properties. Bioinformatics 24:2101-2102. https://doi:10.1093/bioinformatics/btn392.
5. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017)Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722-736. https://doi:10.1101/gr.215087.116.
6. Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34:3094-3100. https://doi:10.1093/bioinformatics/bty191.
7. Maerkens A, Olivé M, Schreiner A, Feldkirchner S, Schessl J, Uszkoreit J, Barkovits K, Güttsches AK, Theis V, Eisenacher M, Tegenthoff M, Goldfarb LG, Schröder R, Schoser B, van der Ven PF, Fürst DO, Vorgerd M, Marcus K, Kley RA (2016) New insights into the protein aggregation pathology in myotilinopathy by combined proteomic and

immunolocalization analyses. Acta Neuropathol Commun 4:8. https://doi:10.1186/s40478-016-0280-0.

8. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575. https://doi:10.1086/519795.
9. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. (2011) Integrative Genomics Viewer. Nature Biotechnology 29:24-26. https://doi:10.1038/nbt.1754.