Protein signature of human white blood cells enables biochemical tracking of rare neurological diseases

Andreas Hentschel (andreas.hentschel@isas.de)
Institute for Analytical Sciences: Leibniz-Institut fur Analytische Wissenschaften - ISAS eV
https://orcid.org/0000-0002-0272-9220
Nina Rademacher
University Hospital Essen Department of Paediatrics III: Universitatsklinikum Essen Klinik fur Kinderheilkunde III
Ulrike Schara-Schmidt
University Hospital Essen: Universitatsklinikum Essen
Nancy Meyer
University Hospital Essen: Universitatsklinikum Essen
Albert Sickmann
Leibniz-Institut für Analytische Wissenschaften – ISAS eV: Leibniz-Institut fur Analytische Wissenschaften - ISAS eV
Anne-Katrin Güttsches
BG University Hospital Bergmannsheil: Berufsgenossenschaftliches Universitatsklinikum Bergmannsheil
Matthias Vorgerd
BG University Hospital Bergmannsheil: Berufsgenossenschaftliches Universitatsklinikum Bergmannsheil
David Schorling
University Medical Center Freiburg: Universitatsklinikum Freiburg
Andreas Roos
BG University Hospital Bergmannsheil: Berufsgenossenschaftliches Universitatsklinikum Bergmannsheil

---

Research Article

Keywords: CHKB, Choline/ethanolamine kinase, Muscular dystrophy, congenital, megaconial type (MDCMC), VAPB, Stathmin

Posted Date: May 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1669058/v1

Abstract

Background: The identification of pathomechanisms leading or contributing to the clinical manifestation of rare neurological diseases such as neuromuscular diseases (NMD) is crucial. The study of the molecular basis of these diseases is also important for the definition of starting points for (new) therapeutic intervention concepts as well as for testing of pathogenicity of genetic variants identified by sequencing approaches. However, these studies are frequently limited by the availability of human biomaterial.

Methods: Human white blood cells were used to generate a spectral library using pH8-fractionation followed by nano liquid chromatography coupled to tandem mass spectrometry. To demonstrate the suitability of this approach, we next subjected proteins isolated from white blood cells derived from two CHKB patients to a data independent acquisition approach. Paradigmatic proteomic findings were confirmed by immunofluorescence in a muscle biopsy derived from one of these patients.

Results: To investigate if human white blood cells might serve as valuable biomaterial for (molecular) studies of NMD, we generated a protein library cataloguing 7543 proteins covering 52.2 % of proteins for which mutations in corresponding genes are known to be causative for NMDs. Based on these findings, we aimed to further demonstrate the suitability of this *in vitro* model to study the etiopathology of a rare NMD caused by bi-allelic variants within *CHKB*. Dysregulation of paradigmatic proteins could be confirmed in the quadriceps muscle biopsy of one of these patients and protein-functions provided new insights into the underlying pathophysiology indicating affection of subcellular structures (functionally) connected to mitochondria such as the Endoplasmic Reticulum (ER) and cytoskeleton. Moreover, altered protein clearance is indicated by our combined proteomic and immunofluorescence findings.

Conclusions: Our combined data reveal that human white blood cells may serve as an *in vitro* system to study the molecular etiology of rare neurological diseases exemplified on a rare mitochondrial NMD based on pathogenic variants affecting *CHKB* in an unbiased fashion.

Introduction

Neuromuscular disorders encompass a spectrum of diseases in which the primary abnormality or lesion is located in the anterior horn cell, the peripheral nerve, the neuromuscular junction or the skeletal muscle. Compared to more common diseases, substantial progress regarding early detection, diagnosis and treatment was made only in selected neuromuscular diseases. Conditions associated with those disorders are usually incurable or difficult to address therapeutically. Very often these diseases present with an early age of onset as well as premature mortality or chronic limitations in mobility and independence. As such, they pose an increasing challenge to families, educational and health as well as care systems, and society in general (e.g. (1–3)). Frequently, these disorders are genetically caused and thus result in a risk of recurrence.

Despite the discovery of numerous genes responsible for many forms of neurological disorders, significant progress has been scarced to date in their biochemical understanding leading to the invention of new therapeutic treatment strategies targeting pathophysiological cascades of these diseases that do not rely on gene replacement. Moreover, even for neuromuscular disorders that are treatable by gene replacement, additional therapeutic approaches are required that address the pathological changes that cannot be reversed by gene/protein restoration (e.g., altered signaling modules) (4). One way to address this issue is to classify diseases into sub-groups based on the major underlying pathomechanism(s), which would lead to the development of common treatment strategies. However, to achieve this, a comprehensive understanding of these systems is mandatory (5).

In the last decade, clinical proteomics has become very important and is increasingly used to study neurological diseases, with a focus on common neurological disorders such as dementia and Parkinson's disease, which pose a major challenge to the health care system and the aging process.

Often, biochemical and functional studies of neurological disorders are ”limited“ to animal models. The main reason for this is the limited availability of human sample material, including suitable human *in vitro* models (6).

Tissue-based molecular diagnostics has enabled great progress on the path to personalized medicine but harbors some significant limitations. The amount of tissue available for analysis is often very limited, especially in the area of rare (but severe) neuromuscular diseases. In addition, tissue biopsies can only provide a snapshot of the disease at a given point in time. Because they are invasive, regular repeated biopsies are not really an option. Thus, from the perspective of utility, tissue biopsies cannot meet the need for serial and long-term monitoring which is required to continuously assess the biochemical background of disease progression and make informed treatment decisions in real time.

Liquid biopsies in combination with clinical mass spectrometry offer the possibility of identifying actively secreted or passively released skeletal muscle and neuronal proteins in the context of pathophysiological processes underlying the respective neuromuscular diseases (7, 8). In this way, new biomarker candidates can be explored to develop improved diagnostic approaches, to design new prognostic tools and to create more reliable methods for the systematic evaluation of experimental/ pre-clinical therapies toward the treatment of neuromuscular diseases in future.

To determine whether white blood cells are suitable for investigating the pathophysiological mechanisms of neuromuscular diseases from a different perspective, different studies including clinical proteomics have already been performed and results suggested the suitability of white blood cells to serve as *in vitro* system enabling the tracking of disease-related processes and to identify modifying factors (e.g. (9–11)). To further elaborate on these findings, in the study presented, a protein/spectral library of human white blood cells was generated and *in silico* signaling pathway analyses were performed to determine the abundance of proteins critical for neuron and muscle development, maintenance, and function. In addition, a comparison of our library with proteins responsible for genetic forms of neuromuscular diseases revealed the expression of a variety of such proteins associated with motor neuron diseases,

congenital myasthenic syndromes, neuropathies/peripheral nervous system diseases, and muscle diseases. Using this library as the basis for data-independent discovery approaches, an exemplary rare inherited neurological disease caused by recessive variants in *CHKB* was investigated. The combined results of proteomic and validation studies provided insights into the underlying pathophysiology showing an affection of subcellular structures closely connected to mitochondria such as the ER and cytoskeleton apart from altered proteolysis. Our combined data therefore support the concept that white blood cells are a valuable model for studying the molecular genesis of neurological diseases.

Results

Protein cataloging of human white blood cells

To elucidate if white blood cells, as part of the liquid biopsy, in combination with clinical mass spectrometry, may serve as a suitable *in vitro* model to study the molecular etiology of neurological conditions such as neuromuscular disorders, and thus help to overcome the limitation of human material for these studies, we generated and analyzed a protein database based on the protein composition of the samples used to generate the protein catalogue (Fig. 1A, Tab. S1). After LC-MS/MS analysis of all fractionated tryptic peptides, this database/ protein catalogue contained 7543 proteins (based on a total number of 81854 peptides). In total, a range of 5 orders of magnitude was covered. The top 5 most abundant peptides identified are part of 5 different proteins including the F-box only protein 6 (binds sugar chains on unfolded glycoproteins promoting their ubiquitination and subsequent degradation), Matrix metalloproteinase-9 (Matrix metalloproteinase that plays an essential role in local proteolysis of the extracellular matrix and in leukocyte migration), Annexin A3, Coronin-1C (required for normal organization of the cytoskeleton, normal cell proliferation and mitochondria distribution) and Pyruvate kinase PKM (catalyzes the transfer of a phosphoryl group from phosphoenolpyruvate (PEP) to ADP, generating ATP). On the other hand, the top 5 lowest abundant proteins that we could identify in our database include mitochondrial protein C2orf69 (involved in respiratory chain, defects in C2orf69 are the cause of a multisystem mitochondrial disorder characterized by auto inflammation, muscular weakness, eye abnormalities, liver dysfunction, epilepsy and cardiomyopathy), HEAT repeat containing protein 5A, Vasodilator-stimulated phosphoprotein (promotes actin elongation and regulates actin dynamics in platelets; important for platelet aggregation), Mixed lineage kinase domain-like protein (pseudo kinase that plays a key role in TNF-induced necroptosis, a programmed cell death process) and Spectrin beta chain, non-erythrocytic 1 (involved in secretion, interacts with calmodulin in a calcium-dependent manner and is thus candidate for the calcium-dependent movement of the cytoskeleton at the membrane).

These 7543 proteins cover all subcellular compartments (Fig. 1B). For the cytosol and the nucleus (including the nucleoplasm) 1076 and 1375 proteins are annotated to be exclusively present in these subcellular compartments, respectively (Fig. 1B). A total of 635 proteins were identified to be localized to mitochondria. The number of identified cell membrane proteins also include such localizing to the endoplasmic reticulum (ER) membrane, mitochondrial membrane, Golgi membrane and nuclear membrane, with 440 proteins only being membrane specific (Fig. 1B). With regard to cellular organelles

such as the endoplasmic reticulum (ER), the Golgi apparatus (Golgi), the lysosome, the peroxisome and the endosomes, we identified 560, 370, 157, 61 and 104 proteins specifically localized to each compartment, respectively (Fig. 1B). The largest overlap was observed for proteins of the ER and the Golgi (Fig. 1B).

To evaluate the proteomic signature of human white blood cells for potential analyses of genetically caused NMD, catalogued proteins were further analysed as follows: proteins were filtered for defects in the corresponding genes that are linked to the manifestation of diseases along the neuromuscular axis including motoneuron, peripheral nervous system diseases, defects in neuromuscular transmission and muscle disorders (https://neuromuscular.wustl.edu/) (12). This approach revealed the expression of 211 from 379 proteins (55.7%) linked to the manifestation of a respective disorder: for hereditary motoneuron disorders (consisting of spinal muscular atrophy (SMA), distal SMA and hereditary moto-neuropathies (HMN)), 40 out of 53 registered proteins (75.5%) were identified. For hereditary motor sensory neuropathies (HMSN), comprising a list of 84 known genes/proteins, 60 (71.4%) were identified including SH3TC2 and GAN (Fig. 1C). In addition, 13 out of 26 known genes/proteins (50.0%) causative for hereditary sensory and autonomic neuropathies (HSAN) were detected in white blood cells. For congenital myasthenic syndrome (CMS) 11 of the currently known causative 35 genes/proteins (31.4%) were identified in our protein/spectral library (Fig. 1C). In addition, from a further list of 238 proteins related to muscle diseases, comprising not only such based on the above-mentioned sources but also candidates based on experimental findings for different muscle diseases, 118 proteins (49.6%) could be identified. These proteins are listed in supplementary tableS2. The spectral library data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD033604.

Global proteomic profiling of CHKB patient derived white blood cells

Having demonstrated the expression of a variety of proteins with neuromuscular significance in human white blood cells, we next aimed to systematically investigate the suitability of these cells for studying the molecular etiology of neuromuscular diseases. To achieve this goal, we investigated white blood cells from patients suffering from a neuromuscular disease associated with recessive variants in the *CHKB* gene. Investigating the proteomic signature of white blood cells, our approach allowed the quantification of 3048 proteins out of which 132 (4.33%) were significantly increased and 157 (5.15%) significantly decreased (Fig. 2A-C, Suppl. Tab. S3). The proteomic profiling data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD033604. To unravel underlying pathomechanisms, *in silico*-based pathway analyses were performed including GO-term and Proteomaps analysis considering the increased and decreased proteins separately.

Cellular processes influenced by proteins with decreased abundance

GO term analysis focused on proteins present at reduced levels in patient-derived white blood cells and revealed the susceptibility of proteins involved in oxidative phosphorylation, modulation of the tricarboxylic acid cycle, ER to Golgi vesicle-mediated transport along with general vesicle-mediated transport, mitochondrial electron transport, (ubiquinol to cytochrome C), mitochondrial electron transport (cytochrome C to oxygen), microtubule cytoskeleton organization and chaperone-mediated protein folding (Fig. 2D). Proteomaps-based analyses revealed alterations in RNA-processing and general metabolism as well as in cytoskeleton and endocytosis (Fig. 2E).

Cellular processes influenced by proteins with increased abundance

Focusing on biological processes, molecular functions, and cellular compartments, our GO term-based analysis revealed that proteins that are more abundant in patient-derived cells include cytoplasmic translation along with RNA-processing, proteolysis, cellular oxidant detoxification along with oxygen transport and response to oxidative stress, actin filament bundle assembly and synaptic vesicle budding from presynaptic endocytic zone membrane (Fig. 2D). Proteomaps-based analyses revealed alterations in cell cycle, amino acid and carbohydrate metabolism as well as in different signaling cascades including RAS and MAPK signaling (Fig. 2E).

Validation of proteomic findings in quadriceps muscle by immunostaining

To validate our proteomic findings, immunostaining studies were carried out on a quadriceps muscle biopsy for paradigmatic proteins identified to be dysregulated in the proteomic signature of white blood cells derived from two *CHKB* patients. To determine their potential dysregulation in the context of mitochondrial pathology, co-staining was performed for all these proteins with Tim23, a mitochondrial import inner membrane translocase subunit. Results of these studies showed that Stathmin is more focally increased within the sarcoplasm showing a co-localization with Tim23 in patient derived muscle fibres whereas in muscle cells of the control the distribution and co-localization is more detected in the sub-sarcolemmal regions (Fig. 3A). Immunostaining of beta-Tubulin revealed a marked reduction in patient derived muscle fibres compared to muscle fibres of the control biopsy (Fig. 3B). However, in some *CHKB* mutant fibres, a focal co-localization of beta-Tubulin and Tim23 was detected within the sarcoplasm (Fig. 3B). Staining of VAPB revealed an immunoreactivity at capillaries and the perimyonuclear region in control fibres whereas in the patient derived biopsy in some fibres a sub-sarcolemmal enrichment was additionally observed (Fig. 3C). Staining of CD63 revealed a co-immunoreactivity with Tim23 at sub-sarcolemmal regions in control fibres, whereas in muscle fibres derived from the patient, this co-localization is restricted to individual fibres with pronounced sarcoplasmic enrichment (Fig. 3D). LAMP2 shows increased co-immunoreactivity with Tim23 within the sarcoplasm of *CHKB* mutant fibres (Fig. 3E). For Cytochrome C, a reduced immunoreactivity was observed within muscle fibres of the *CHKB* patient compared to the matched control (Fig. 3F).

Discussion

Protein cataloguing in human white blood cells reveals expression of a variety of proteins of neurological relevance

Due to the progressive degradation and damage of muscle fibres, as well as the dissolution and regression of motor neurons and the myelin sheath, protein leakage into the bloodstream occurs and indeed related proteins such as muscular creatine kinase and neurofilaments became validated biomarkers mirroring muscle damage in disease degeneration measurable in blood, respectively. However, these proteins are “only” secondarily linked to the pathophysiological processes and rather represent non-specific markers. In the pre-screening of candidate genes as well as in the evaluation of genetic variants of unknown significance, analyses of protein extracts from muscle and/ or nerve biopsies are performed as a procedure in the diagnostic work-up of patients suffering from neuromuscular diseases (5, 13). However, collection of this biomaterial is invasive and suitable alternatives would be beneficial for the management of patients. Here, we address the suitability of white blood cells collected in terms of a liquid biopsy to serve as an alternative of the more invasive muscle and nerve biopsies. Indeed, protein cataloguing revealed that 55.7% of proteins of neuromuscular relevance are expressed in these cells. This in turn highlights that for more than half of the known NME-relevant proteins, a biochemical monitoring (including the evaluation of ambiguous genetic variants on protein abundance and the study of the effect of loss or gain of protein function on downstream cascades) would be possible in white cells enabling new avenues in the diagnostic management of the patients as well as in the definition of new starting points for therapeutic intervention concepts. The finding that the coverage of protein encoded by genes responsible for myopathic disorders is much higher than the ones for diseases affecting the nervous system (Fig. 1C) may be related to the fact that muscle and white blood cells derived from the same embryonic layer, the mesoderm. In contrast, neuronal cells are derived from the ectoderm bearing the neuroectoderm.

Focusing on individual proteins, it is worth noting that SH3TC2 and GAN were identified to be expressed on the protein level in white blood cells. Interestingly, SH3TC2 expression was thought to be limited to Schwann cells (14, 15)). Both proteins are encoded by genes known to be responsible for recessive forms of peripheral neuropathies. This might open new avenues in (i) the diagnostic work-up of patients with clinical suspicion of these both subtypes of neuropathy, making a nerve biopsy less essential as well as in (ii) the study of the precise molecular nature of this disease group and (iii) to pre-clinically test therapeutic concepts such as gene therapy which has already been studied in *Sh3tc2*-deficient mice (16).

Taking into consideration that the generation of our spectral library was mostly based on white blood cells derived from healthy controls and patients showing a vulnerability of the peripheral nervous system, results of our proteomic profiling on *CHKB* patient derived white blood cells highlight the suitability of this spectral library also to decipher biochemical changes related to muscular diseases. This assumption is supported by the results of our validation studies on a muscle biopsy (see below).

Proteomic studies on white blood cells of a CHKB **patients allowed insights into the molecular etiology of the disease**

CHKB is a rate-limiting enzyme in the synthesis of phosphatidylcholine, a predominant mitochondrial membrane phospholipid. Pathogenic variants in the corresponding *CHKB* gene were linked to the manifestation of megaconial congenital muscular dystrophy (CMD; OMIM #602541), a rare autosomal recessive disorder. To date, only 40 confirmed patients are recorded (17). The muscular dystrophy is characterized by early-onset muscle wasting and is complicated by mental retardation, and dilated cardiomyopathy in half of affected individuals. Some patients may die from cardiomyopathy in the first or second decade of life. Muscle biopsy shows peculiar, enlarged mitochondria that are prevalent toward the periphery of the fibres but are sparse in the center (18). Based on the known expression of CHKB in white blood cells as highlighted by our protein catalogue, we investigated the proteomic signature of white blood cells obtained from two genetically confirmed *CHKB* patients identifying the significant dysregulation of a total of 289 proteins (9.48% of all quantified proteins). This approach along with subsequent *in silico* and immunological-based confirmational studies on the muscle biopsy of one of these *CHKB* patients warranted indications toward further insights into underlying pathophysiological processes: altered mitochondrial function is indicated by the results of our GO-term based studies in turn suggesting the pathogenicity of the identified variants within *CHKB*. Our validation immunofluorescence studies confirmed reduced abundance of Cytochrome C thus supporting the proteomic findings obtained in white blood cells. Along this line, our immunofluorescence findings did not only confirm an increase of CD63 and LAMP2, two marker proteins of autophagic protein clearance, which have also been identified in white blood cells by proteomic profiling but moreover showed a strong co-localization with Tim23, a mitochondrial protein, within the sarcoplasm thus suggesting mitophagy. This assumption is supported by a pervious study of Mitsuhashi and Nishino demonstrating that a phosphatidylcholine biosynthetic defect leads to mitochondrial dysfunction and increased mitophagy in the disease cause of *CHKB* (19). Mitochondria form close physical associations with the Endoplasmic Reticulum (ER) that regulate a number of physiological functions modulated by VAPB. Hereby VAPB acts a scaffold to tether the two organelles and to modulate autophagy. Depletion of VAPB expression stimulates autophagosome formation (20). Decreased abundance of VAPB in white blood cells along with the irregular sub-sarcolemmal localization in a proportion of *CHKB* mutant muscle fibres might accord with the concept of altered mitophagy. However, further functional studies on *CHKB* mutant cells are needed to decipher a potential molecular interplay of VABP and mitophagy in the context of this disease. Nevertheless, one might postulate that (immortalized) white blood cells represent a suitable *in vitro* model to further study and modulate mitophagy in *CHKB*-related pathology, an important aspect in the definition and testing of therapeutic interventions.

Moreover, our data indicate altered cytoskeleton by dysregulation of different associated proteins including Tubulins. Tubulin has been described as an inherent component of mitochondrial membranes interacting with the voltage-dependent anion channel as the main component of the permeability transition pore (21). Of note, the voltage-dependent anion-selective channel protein 2 is significantly downregulated in *CHKB* patient derived white blood cells as shown by our proteomic findings. Hence, one

might assume that loss of functional CHKB causes a pathophysiological cross-talk of tubulins and the voltage-dependent anion channel by affecting their respective abundances. Indeed, in muscle cells derived from our *CHKB* patient two, generalized decrease of beta-Tubulin expression was identified along with the presence of sarcoplasmic dots immunoreactive for Tim23 and beta-Tubulin. Of note, Stathmin is increased in white blood cells as identified by our proteomic study and shows a focal sarcoplasmic co-immunoreactivity with Tim23 in muscle cells of *CHKB* patient two. Aberrant Stathmin increase was already linked to a decreased level of polymerized Tubulin, correlated with disease severity in a mouse model of 5q-associated (SMN-related) of spinal muscular atrophy. A pathophysiological cross-talk in this muscle wasting condition was moreover supported by the finding of reduced microtubule densities and beta-Tubulin levels in distal axons of affected SMA-like mice and an impaired microtubule network in Smn-deficient cells (22). However, further functional studies are needed to confirm the hypothesis of a pathophysiological cross-talk of Tubulins, Stathmin and diseased mitochondria in the *CHKB*-etiopathology.

Conclusions

- Human white blood cells express 211 proteins of neuromuscular relevance (57.7% of all currently known proteins linked to this disease entity) thus declaring this cellular population as an *in vitro* model enabling to track associated pathophysiological processes and to test potential treatment strategies pre-clinically.
- Proteomic profiling of white blood cells derived from *CHKB* patients allowed the identification of protein dysregulation also showing a relevance in skeletal muscle as a primary vulnerable tissue of the disease. Therefore, this cellular population might be utilized to track disease processes of neuromuscular disorders.
- Altered mitophagy and Tubulin-pathology seem to be part of the pathophysiological spectrum associated with the presence of pathogenic *CHKB* variants.

Patients, Materials & Methods

Patient information

CHKB case 1:

This is a now 3-year old girl. She was born preterm at the age of 30 weeks of gestation due to patholgogical CTG after otherwise unremarkable pregnancy. Postnatal, she required non-invasive respiratory support for some days. She showed severely delayed motor development, with unability to roll from supine to prone at age of 12 months. Her cMRI showed a rather thin corpus callosum and mild white matter abnormalities. Creatinine kinase values were elevated (max. 1350 U/l). Cardiomyopathy was ruled out by echocardiography. Genetic workup included karyotyping, array-CGH and exome analysis. The latter revealed the compound heterozygous mutations in *CHKB* [c.869dupA(p.His290fs) and c.419delC(p.Pro140fs)], leading to a frame-shift and a premature stop-codon. Both parents were found to

carry one of the mutations. Now, at the age of 3 years she has little head control, very reduced spontaneous movements and is not able to sit or crawl.

CHKB patient 2:

This boy was born full term to healthy first-degree consanguine parents of Arabian descent. Up to the age of 16 months, the parents describe a normal development. At the age of 1 year and 11 months, he was first seen in the clinic of pediatric neurology due to persistently elevated creatine kinase (206 U/l-maximum of 4096 U/l). Leading symptoms on neuropediatric examination were weakness of the trunk and proximal limb muscles, a general muscle hypotonia and a mild pseudohypertrophy of thighs and calves. He could walk independently with wide base, was holding onto handrail when climbing stairs and showed a Gowers's sign. In addition, he showed severe generalized developmental delay with hyperactive behavior and stereotyped hand movements. A muscle biopsy was performed at 4 years and 11 months. Histochemical and biochemical analyses hinted towards a mitochondriopathy. Genetic workup, including a microarray-analysis, a screening for mitochondrial (mtDNA) depletion syndromes and complete sequencing of mtDNA, were without pathological findings. A multi-gene panel analysis revealed a novel homozygous variant c.248.dup (p.Arg84Profs*126) in exon 2 within the *CHKB* gene leading to a loss-of function.

Preparation of leukocytes for mass spectrometry and spectral library generation

Two pools of white blood cells were created for spectral library generation: the first pool containing white blood cells from healthy donors (8 samples) and the second pool containing white blood cells derived from patients suffering from different forms of neuromuscular diseases (6 samples) (Tab. 1). Each pool was processed as follows:

The pooled snap-frozen white blood cells were lysed in a buffer containing 5% SDS, 50 mM Tris, 150 mM NaCl, pH 7.8 and cOmplete™ ULTRA protease inhibitor using the Bioruptor® (Diagenode) for 10 min (30 seconds on, 30 seconds off, 10 cycles) at 4 °C. Subsequently, 20 µl of each sample was collected and diluted 1:4 with 10 mM ammonium bicarbonate buffer, pH 7.8 (ABC) to perform a BCA-based determination of protein concentration according to the manufacturer's instructions (Pierce BCA protein assay kit). Reduction and carbamidomethylation of the remaining samples was performed utilizing 10 mM tris-(2-carboxyethyl)-phosphine (TCEP) for 30 min at 37 °C, followed by application of 15 mM iodoacetamide (IAA) for an additional 30 min at room temperature (RT) in the dark.

Samples were further processed using the S-Trap™ (Protifi) sample preparation procedure: after acidifying the samples by adding 12% aqueous phosphoric acid, they were diluted with S-Trap binding buffer (90% methanol (MeOH), triethylammonium bicarbonate (TEAB) 100 mM, pH 7.1). Loading of proteins onto the S-trap columns, including centrifugation steps, was performed according to the manufacturer's instructions. Filter-based tryptic digestion was performed for 2 hours at 47 °C with a trypsin to protein ratio of 1:20. Peptides were then eluted in several elution steps, starting with 10 mM ABC, followed by elution with 0.1% formic acid (FA), and finally with 80% acetonitrile (ACN). Eluted peptides were dried in a

vacuum concentrator and then dissolved in 0.1% TFA for subsequent LC-MS/MS analysis or in 10 mM ammonium acetate containing 0.4 mM FA (pH 8.0) for subsequent pH8 reversed-phase fractionation.

To proof the efficacy of the tryptic digest, proteolytic lysates were analyzed using a monolithic column separation system (PepSwift monolithic PS-DVB PL-CAP200-PM, Dionex) on an inert Ultimate 3000 HPLC (Dionex) by direct injection of 1 µg sample. A binary gradient (solvent A: 0.1% TFA, solvent B: 0.08% TFA, 84% ACN) of 5-12% B in 5 min and then of 12-50% B in 15 min at a flow rate of 2.2 µl/min and 60 °C was used. UV traces were recorded at 214 nm (23).

Table 1: List of controls and patients used to generate the spectral library. 1st pool included 4 male and 4 female controls while 2nd pool comprised 6 samples from patients.

Nr.	Sex	Age	Diagnosis
1	Male controls	45	None
2		41	None
3		35	None
4		26	None
1	Female controls	34	None
2		35	None
3		56	None
4		38	None
1	Neurological patients	15	Severe idiopathic neuropathy
2		18	Superordinate idiopathic neurologic disease
3		12	Superordinate idiopathic neurologic disease
4		8	Perinatal intracerebral haemorrhage: *de novo* variant in *COL4A1*
5		18	Severe idiopathic neuropathy
6		16	Severe idiopathic neuropathy

PH8-based sample fractionation

Each of the digested and desalted samples selected for subsequent spectral library generation was first dried in a vacuum concentrator. Peptides were then dissolved in a buffer containing 10 mM ammonium

acetate and 0.4 mM formate (pH 8.0) (concentration 50µg/µl) and separated on a C18-RP chromatography column (loading amount 50 µg). Peptides were loaded onto the column with solvent A (10 mM ammonium acetate, 0.4 mM formate, pH 8.0) at a flow rate of 12.5 µl/min. Separation and fractionation were performed with the following gradient using solvent B (84% acetonitrile in 10 mM ammonium acetate, 0.4 mM formate, pH 8.0): 3-10% in 10 min, 10-25% for 35 min, 25-40% for 20 min, 40-95% for 10 min, 95% for 5 min, and 20 min equilibration at 3%. The individual fractions were collected at an interval of 60 sec, and each sample was divided into 15 fractions. The collection was done in a time interval from 10 to 75 min of the gradient. The fractions were collected in a concatenated approach. After fractionation, individual samples were allowed to dry in a vacuum concentrator and dissolved in 0.1% TFA prior to subsequent nano-LC-MS/MS analysis (1 µg/µl).

Generating the spectral library

Since the establishment of a spectral library is a requirement for performing data-independent LC-MS/MS-based sample analysis, all fractions obtained from the pH8 fractionation mentioned previously were analyzed by nano-LC-MS/MS using 1 µg of each: Samples were loaded onto an Ultimate 3000 Rapid Separation Liquid Chromatography (RSLC) nano system ProFlow flow controlled and coupled to a Fusion Lumos Tribrid mass spectrometer (both from Thermo Scientific). After loading, peptides were concentrated on a trapping column (Acclaim C18 PepMap100, 100 µm, 2 cm) using 0.1% TFA at a flow rate of 10 µl/min. Subsequent sample separation was performed on a reverse phase column (Acclaim C18 PepMap100, 75 µm, 50 cm) using a linear binary gradient: 3% solvent B (84% ACN with 0.1% TFA) for 10 min, a linear increase in solvent B to 35% for 120 min, a linear increase in solvent B to 95% for 10 min, followed by a linear decrease in solvent B to 3% for 5 min. MS survey scans were acquired on the Fusion Lumos using the following settings: The mass spectrometer was operated in data-dependent acquisition (DDA) mode with full MS scans from 300 to 1500 m/z at a resolution of 120,000 (Orbitrap) using the polysiloxane ion at 445.12002 m/z as lock mass. The automatic gain control (AGC) was set to 2E5, and the maximum injection time was set to 50 milliseconds. The most intense ions above a threshold of 5E3 were selected for fragmentation at a normalized collision energy (nCE) of 30% (HCD) in each cycle of acquisition analysis after each survey scan. The dynamic exclusion time was set to 15 sec. The number of precursor ions selected for fragmentation was determined by the "fast" acquisition algorithm. Fragment ions were acquired in the linear ion trap with an AGC of 1E4 and a maximum injection time of 35 milliseconds.

The data acquired was imported into Spectronaut software (Biognosys). The human proteome data from UniProt (www.uniprot.org), containing 20,374 entries, were selected as the proteome background. Processing settings were set as follows: the enzyme was trypsin, the minimum and maximum peptide lengths were set to 7 and 52, respectively, and missed cleavages were set to 2. Carbamidomethyl for cysteine was set as a fixed modification, and acetyl (protein N-term) and oxidation of methionine were set as variable modifications. All library generation settings, including tolerances, identification, filters, iRT calibration, and workflow, were set to factory defaults. For relative quantification, the Top N max 3 option was selected.

Data independent acquisition (DIA-LC-MS/MS)

For data-independent acquisition (DIA), an identical nano-LC-MS/MS setup was used as for the DDA acquisition. 1 µg of each sample was subjected to analysis and mixed with an appropriate amount of standard iRT peptides (Biognosys). Full MS scans were acquired from 300-1100 m/z with a resolution of 60,000 (Orbitrap) using the polysiloxane ion at 445.12002 m/z as the lock mass. The automatic gain control (AGC) was set to 5E5, and the maximum injection time was set to 20 milliseconds. The full MS scans were followed by 30 DIA windows acquired at a resolution of 30,000 (Orbitrap) with an AGC of 1E6, a maximum injection time of 60 milliseconds, and an nCE of 32 (HCD).

Analysis of the DIA data

For the analysis of the samples acquired with nano-LC-MS/MS in DIA mode, the data were submitted to the Spectronaut software and analyzed with a library-based search. The library used was the spectra library created previously. The search and extraction settings were kept as default (BGS factory settings). The human proteome data from UniProt (www.uniprot.org), containing 20,374 entries, were selected as the proteome background.

Immunofluorescence studies

For the validation of proteomic findings, immunofluorescence studies were carried out on quadriceps muscle biopsy specimen derived from *CHKB* patient two as well as a gender- and age-matched control according to a protocol published previously (24). Doing so, the following primary antibodies were used: CD63 (ab8219; 1:200), LAMP2 (MA1-165; 1:200), CDC42 (ab41429; 1:200), Cytochrome C (ab133504; 1:200), alpha-Tubulin (GTX628802; 1:200), Cofilin-1 (ab42824; 1:200); VAPB ((25); 1:200), Stathmin (ab52630; 1:200), Tim23 (bd transduction laboratories 611222; 1:200), anti-rabbit-488 (Invitrogen (a27034); 1:500), anti-mouse-488 (Invitrogen (a28175); 1:500), anti-mouse-594 (Invitrogen (a11032); 1:500).

Declarations

Ethics approval and consent to participate

Informed consent was obtained from all patients and the ethics committee of University Medicine Essen (19-9011-BO) had granted ethical approval.

Consent for publication

Not applicable

Availability of data and materials

The mass spectrometry proteomics data regarding the spectral library and the proteomic profiling data have been deposited to the ProteomeXchange Consortium via the PRIDE (26) partner repository with the

dataset identifier PXD033604.

Access for reviewers:

Username: reviewer_pxd033604@ebi.ac.uk

Password: 91bPEZX9

Competing interests

The authors declare not to have any competing conflict of interest.

Funding

Financial support by the Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen, the Senatsverwaltung für Wirtschaft, Technologie und Forschung des Landes Berlin and the Bundesministerium für Bildung und Forschung is gratefully acknowledged. Parts of this study were financed in the framework of the NME-GPS project (www.nmd-gps.net) by the European Regional Development Fund (ERDF). This work was also supported by a grant from the AFM (to A.R. #21644). AR, MV and USS moreover received funding from the German Society of Muscle Diseases (DGM; Deutsche Gesellschaft für Muskelkranke e.V.) for this project.

Author contributions

AR, AS, USS, AKG and MV designed the study. AH performed the proteomic experiments. NM performed the immunofluorescence-studies. NR and DS investigated the patients and provided their white blood cells. AR and AH drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Drs. Adela Della Marine, Andrea Gangfuss and Heike Kölbel for providing the white blood cell samples utilized to generate the spectral library.

References

1. Landfeldt E, Pechmann A, McMillan HJ, Lochmuller H, Sejersen T. Costs of Illness of Spinal Muscular Atrophy: A Systematic Review. Appl Health Econ Health Policy. 2021;19(4):501–20.
2. Landfeldt E, Nikolenko N, Jimenez-Moreno C, Cumming S, Monckton DG, Gorman G, et al. Disease burden of myotonic dystrophy type 1. J Neurol. 2019;266(4):998–1006.
3. McMillan HJ, Gerber B, Cowling T, Khuu W, Mayer M, Wu JW, et al. Burden of Spinal Muscular Atrophy (SMA) on Patients and Caregivers in Canada. J Neuromuscul Dis. 2021;8(4):553–68.
4. Hensel N, Kubinski S, Claus P. The Need for SMN-Independent Treatments of Spinal Muscular Atrophy (SMA) to Complement SMN-Enhancing Drugs. Front Neurol. 2020;11:45.

5. Thompson R, Spendiff S, Roos A, Bourque PR, Warman Chardon J, Kirschner J, et al. Advances in the diagnosis of inherited neuromuscular diseases and implications for therapy development. Lancet Neurol. 2020;19(6):522–32.
6. Hentschel A, Czech A, Munchberg U, Freier E, Schara-Schmidt U, Sickmann A, et al. Protein signature of human skin fibroblasts allows the study of the molecular etiology of rare neurological diseases. Orphanet J Rare Dis. 2021;16(1):73.
7. Spitali P, Hettne K, Tsonaka R, Charrout M, van den Bergen J, Koeks Z, et al. Tracking disease progression non-invasively in Duchenne and Becker muscular dystrophies. J Cachexia Sarcopenia Muscle. 2018;9(4):715–26.
8. Schorling DC, Kolbel H, Hentschel A, Pechmann A, Meyer N, Wirth B, et al. Cathepsin D as biomarker in cerebrospinal fluid of nusinersen-treated patients with spinal muscular atrophy. Eur J Neurol. 2022.
9. Cox D, Henderson M, Straub V, Barresi R. A simple and rapid immunoassay predicts dysferlinopathies in peripheral blood film. Neuromuscul Disord. 2019;29(11):874–80.
10. Gangfuss A, Hentschel A, Rademacher N, Sickmann A, Stuve B, Horvath R, et al. Identification of a novel homozygous synthesis of cytochrome c oxidase 2 variant in siblings with early-onset axonal Charcot-Marie-Tooth disease. Hum Mutat. 2022;43(4):477–86.
11. Kollipara L, Buchkremer S, Coraspe JAG, Hathazi D, Senderek J, Weis J, et al. In-depth phenotyping of lymphoblastoid cells suggests selective cellular vulnerability in Marinesco-Sjogren syndrome. Oncotarget. 2017;8(40):68493–516.
12. Benarroch L, Bonne G, Rivier F, Hamroun D. The 2020 version of the gene table of neuromuscular disorders (nuclear genome). Neuromuscul Disord. 2019;29(12):980–1018.
13. Roos A, Thompson R, Horvath R, Lochmuller H, Sickmann A. Intersection of Proteomics and Genomics to "Solve the Unsolved" in Rare Disorders such as Neurodegenerative and Neuromuscular Diseases. Proteomics Clin Appl. 2018;12(2).
14. Senderek J, Bergmann C, Stendel C, Kirfel J, Verpoorten N, De Jonghe P, et al. Mutations in a gene encoding a novel SH3/TPR domain protein cause autosomal recessive Charcot-Marie-Tooth type 4C neuropathy. Am J Hum Genet. 2003;73(5):1106–19.
15. Phan V, Schmidt J, Matyash V, Malchow S, Thanisch M, Lorenz C, et al. Characterization of Naive and Vitamin C-Treated Mouse Schwann Cell Line MSC80: Induction of the Antioxidative Thioredoxin Related Transmembrane Protein 1. J Proteome Res. 2018;17(9):2925–36.
16. Schiza N, Georgiou E, Kagiava A, Medard JJ, Richter J, Tryfonos C, et al. Gene replacement therapy in a model of Charcot-Marie-Tooth 4C neuropathy. Brain. 2019;142(5):1227–41.
17. Bardhan M, Polavarapu K, Bevinahalli NN, Veeramani PK, Anjanappa RM, Arunachal G, et al. Megaconial congenital muscular dystrophy secondary to novel CHKB mutations resemble atypical Rett syndrome. J Hum Genet. 2021;66(8):813–23.
18. Mitsuhashi S, Ohkuma A, Talim B, Karahashi M, Koumura T, Aoyama C, et al. A congenital muscular dystrophy with mitochondrial structural abnormalities caused by defective de novo

phosphatidylcholine biosynthesis. Am J Hum Genet. 2011;88(6):845–51.
19. Mitsuhashi S, Nishino I. Phospholipid synthetic defect and mitophagy in muscle disease. Autophagy. 2011;7(12):1559–61.
20. Gomez-Suaga P, Paillusson S, Stoica R, Noble W, Hanger DP, Miller CCJ. The ER-Mitochondria Tethering Complex VAPB-PTPIP51 Regulates Autophagy. Curr Biol. 2017;27(3):371–85.
21. Carre M, Andre N, Carles G, Borghi H, Brichese L, Briand C, et al. Tubulin is an inherent component of mitochondrial membranes that interacts with the voltage-dependent anion channel. J Biol Chem. 2002;277(37):33664–9.
22. Wen HL, Lin YT, Ting CH, Lin-Chao S, Li H, Hsieh-Li HM. Stathmin, a microtubule-destabilizing protein, is dysregulated in spinal muscular atrophy. Hum Mol Genet. 2010;19(9):1766–78.
23. Burkhart JM, Schumbrutzki C, Wortelkamp S, Sickmann A, Zahedi RP. Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. J Proteom. 2012;75(4):1454–62.
24. Roos A, Hathazi D, Schara U. Immunofluorescence-Based Analysis of Caveolin-3 in the Diagnostic Management of Neuromuscular Diseases. Methods Mol Biol. 2020;2169:197–216.
25. Nachreiner T, Esser M, Tenten V, Troost D, Weis J, Kruttgen A. Novel splice variants of the amyotrophic lateral sclerosis-associated gene VAPB expressed in human tissues. Biochem Biophys Res Commun. 2010;394(3):703–8.
26. Perez-Riverol Y, Bai J, Bandla C, Garcia-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 2022;50(D1):D543-D52.

Figures

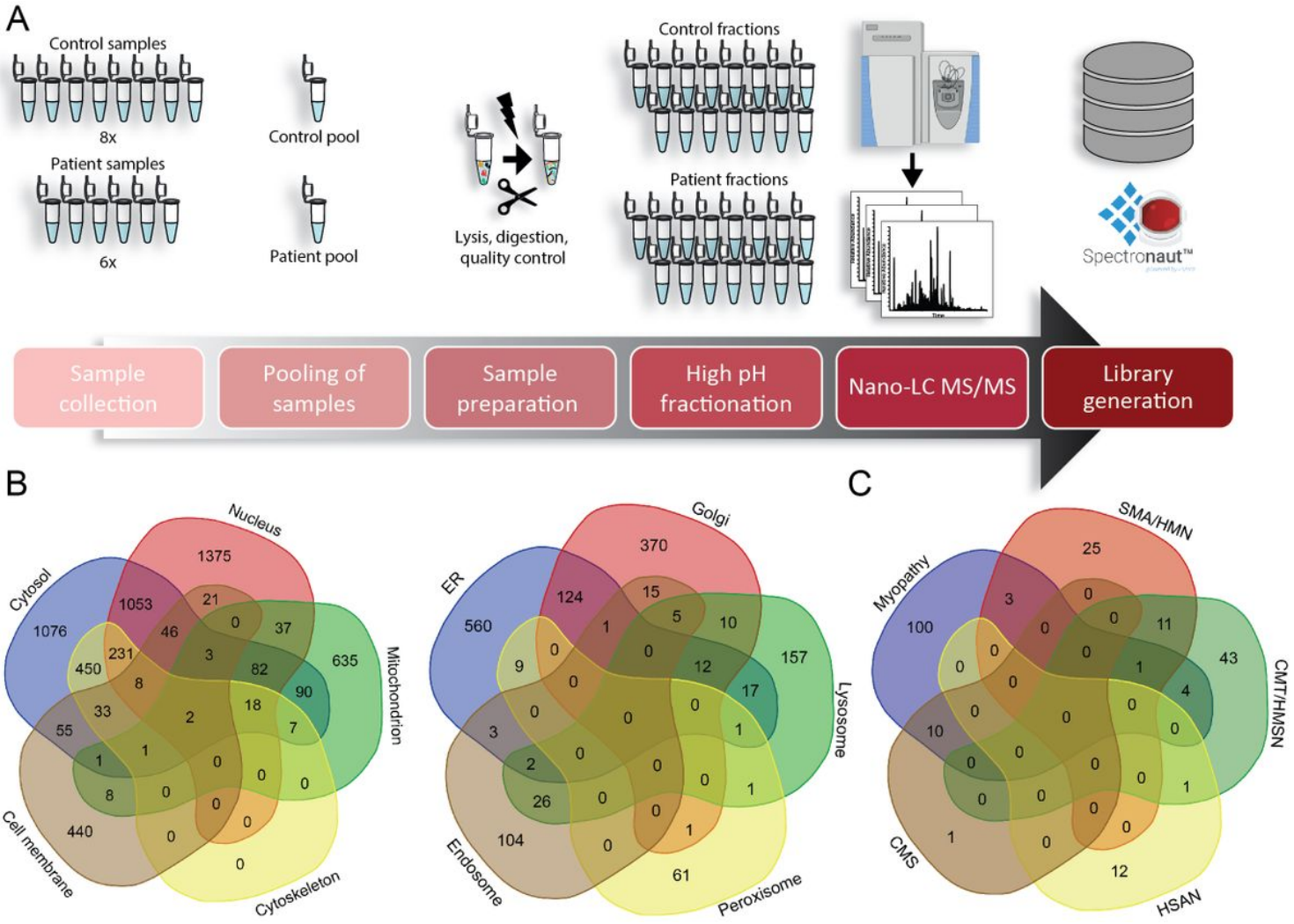

Figure 1

Protein cataloguing in human white blood cells: (A) Schematic representation of the applied workflow towards the generation of a spectral library of human white blood cells (based on 14 samples; eight derived from control individuals and six patients suffering from neurological diseases). (B) Distribution of identified proteins across the different subcellular compartments and organelles. (C) Coverage of different neurological diseases by proteins encoded by genes causative for the manifestation of different subtypes of these diseases.

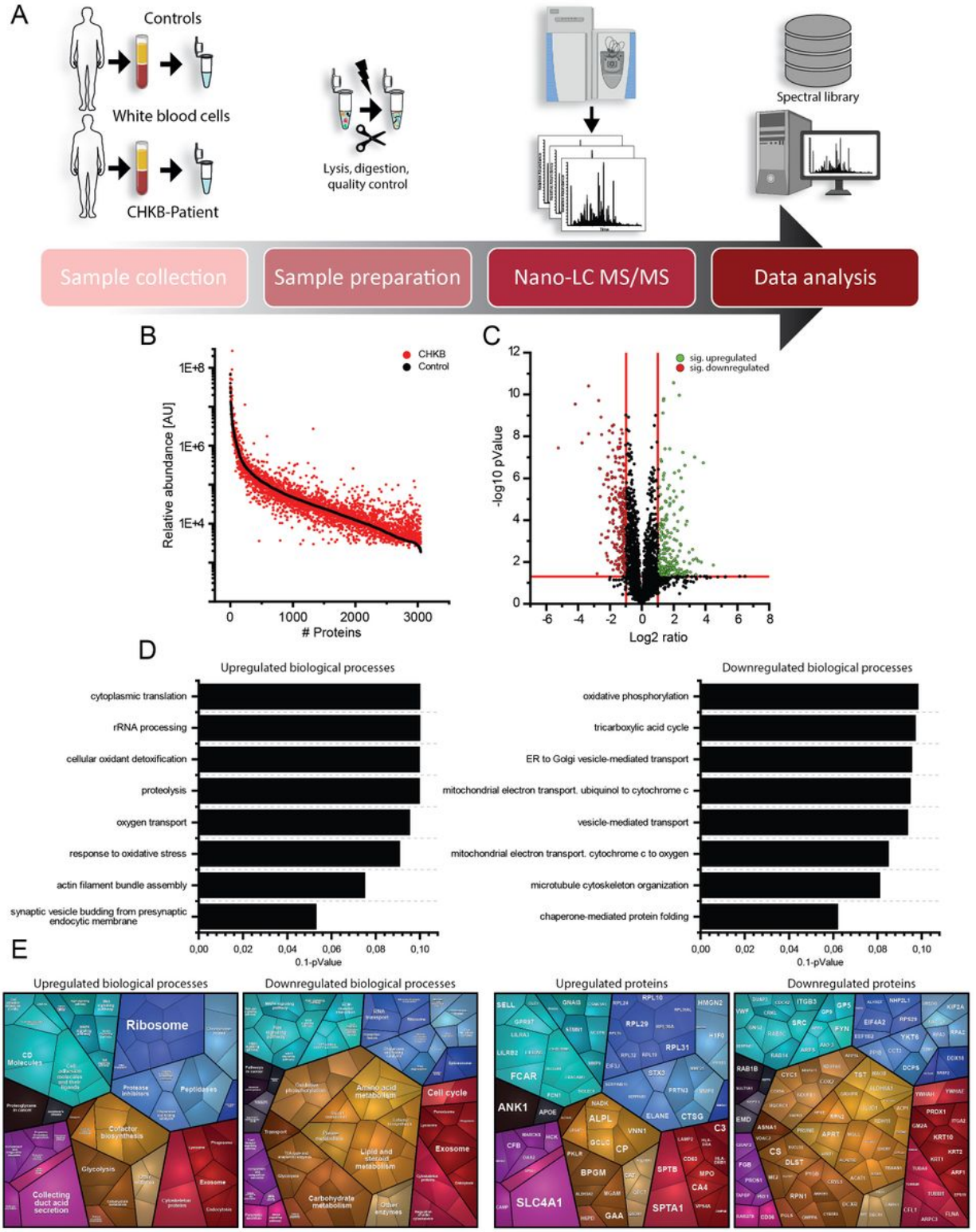

Figure 2

Proteomic profiling of human white blood cells derived from two *CHKB* patients and controls: (A) Schematic representation of the applied workflow toward the identification of protein dysregulations in white blood cells derived from *CHKB* patients. (B) Abundance plot showing the dynamic range of all identified proteins based on their relative quantification using always the 3 highest abundant peptides for each protein, allowing protein comparison within an experiment. All identified proteins of the control

(black) are sorted with decreasing abundance while the patient (red) was plotted in the same order to directly compare the different abundances. (C) Volcano plot highlighting statistically significant increased proteins (green dots) as well as decreased proteins (red dots). (D) GO-term based analysis of biological processes affected by upregulated proteins (left) and downregulated proteins (right). (E) Proteomaps-based analysis of biological processes affected by upregulated proteins and downregulated proteins.

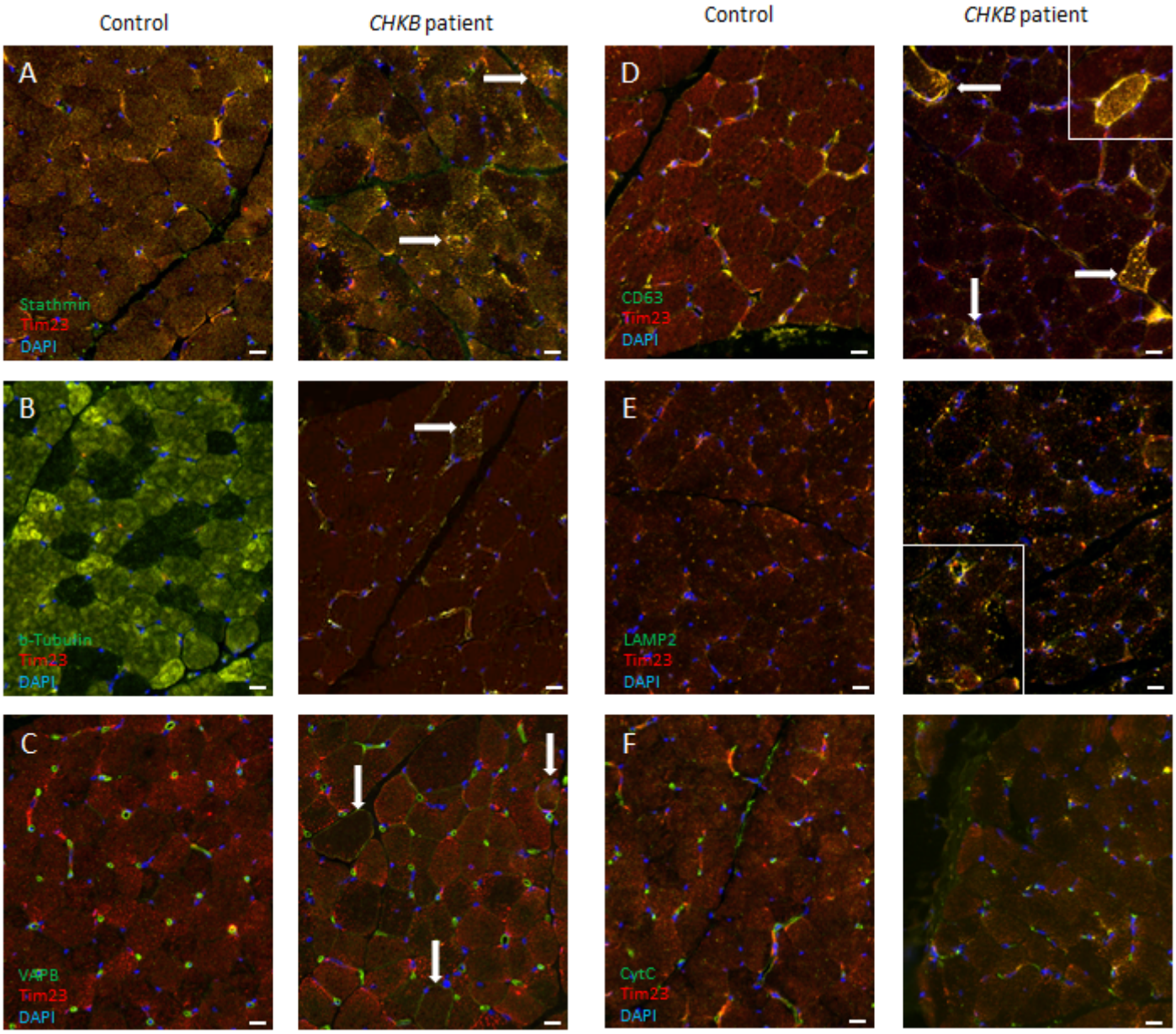

Figure 3

Immunofluorescence studies on a quadriceps muscle biopsy derived from a *CHKB* patient: (A) Co-staining of Stathmin (green) and Tim23 (always displayed red) revealing increased presence of sarcoplasmic dots co-immunoreactive for these proteins (white arrows). (B) Generalized reduced

immunoreactivity of Tubulin (b-Tubulin, green) and presence of sarcoplasmic dots co-immunoreactive for b-Tubulin and Tim23 in *CHKB* mutant muscle cells (white arrows). (C) Presence of sarcolemmal immunoreactivity of VAPB (green) in muscle biopsy derived from the *CHKB* patient (white arrows). (D) Increased co-immunoreactivity of CD63 (green) and Tim23 in a proportion of *CHKB* mutant muscle cells (white arrows; see also insert). (E) Sarcoplasmic increase of dots co-immunoreactive for LAMP2 (green) and Tim23 in *CHKB* mutant muscle cells (see also inset). (F) Generalized reduced immunoreactivity of Cytochrome C (CytC, green) in *CHKB* mutant muscle cells. Nuclei are displayed with DAPI (blue). Scale bars: 30μm

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx
- TableS2.docx
- TableS3.docx