Characterizing protein processing in the Endoplasmic Reticulum using quantitative proteomics: the pathogenesis of the Marinesco-Sjörgren Syndrome

Zur Erlangung des akademischen Grades eines

Dr. rer. nat

von der Fakultät Bio- und Chemieingenieurwesen
der Technischen Universität Dortmund
genehmigte Dissertation

vorgelegt von

MSc. Laxmikanth Kollipara

aus

Hyderabad, India

Tag der mündlichen Prüfung: 20.12.2016

1. Gutachter: Prof. Dr. Albert Sickmann

2. Gutachter: Prof. Dr. Oliver Kayser

Dortmund 2016

1. Prüfer: Prof. Dr. Markus Nett

Abstract

In this work, characterization of Marinesco-Sjögren Syndrome (MSS) was performed for the first time on the proteome level using mass spectrometry (MS)-based quantitative proteomics strategies. MSS is a neuromuscular and neurodegenerative disorder and it is caused due to the mutational inactivation of SIL1 protein, which results in malfunctioning of protein folding machinery mediated by the chaperone BiP that can lead to the ER stress-induced cell death via apoptotic signaling. The major goals were (i) to understand the rescue mechanisms in SIL1-deficient non-vulnerable tissues from human and (ii) to verify the cellular perturbations caused due to the loss of functional SIL1 in *woozy* mouse (i.e. mouse model of MSS). To achieve these aims, samples derived from five different MSS cases and two different tissues from *woozy* along with their respective healthy controls were studied.

For this, comparative LC-MS proteomics approaches such as chemical labeling (i.e. iTRAQ) and label-free quantification (precursor ion intensity based and NSAF) were employed. During which, sample preparation workflows were optimized that enabled to process clinical samples related to MSS that included primary cell lines and mammalian tissues for the subsequent quantitative LC-MS analyses. This also included an investigation of occurrence of artificial protein carbamylation, which is a well-known unwanted artefact in quantitative proteomics.

By employing these workflows, abundances of thousands of proteins in both MSS patients and *woozy* were relatively quantified. Among these, the number of differentially regulated proteins varied depending on the cell/tissue type and the clinical state i.e. SIL1-affected or unaffected. However, in both conditions, the absence of functional SIL1 showed disturbed cellular activities suggesting its important role in BiP-mediated protein folding process. In MSS SIL1 non-vulnerable tissues data, the processes which might mitigate the ER-stress induced due to SIL1 loss were identified. Next, proteome analysis of SIL1 depleted HEK293 cell line was performed to study the pathophysiology of SIL1 loss in more detail. Additionally, a targeted-MS based method was developed to assay proteins that are involved in the unfolded protein response pathway, which is triggered under the ER-stress conditions. Lastly, proteomic profiling of a human myoblastic RCMH cell line was carried out that can serve as an in vitro model to investigate muscle and neuromuscular disorders.

Abstract (in German)

In dieser Arbeit wurde eine Charakterisierung des Marinesco-Sjörgen Syndroms (MSS) zum ersten Mal auf Proteom-Ebene mittels Massenspektrometrie-basierten, quantitativen Methoden durchgeführt. MSS ist eine neuromuskuläre und neurodegenerative Erkrankung und wird verursacht durch mutations-bedingte Inaktivierung des SIL1 Proteins. Diese führt zu einer Störung der Proteinfaltung durch das Chaperonprotein BiP und somit zu einer ER-Stresssituation mit möglichem folgendem Zelltod durch Apoptose. Das Hauptziel bestand darin, die Rettungsmechanismen in SIL1 unabhängigen Geweben (human) zu verstehen, sowie bei fehlendem funktionellen SIL1 in *woozy* Mäusen (MSS-Mausmodell) die verursachten Störungen auf zellulärer Ebene zu bestimmen.

Hierzu wurden vergleichende, proteomische Strategien, wie chemische Isotopenmarkierung und Label-freie Quantifizierung angewandt. Innerhalb der Durchführung wurden Methoden der Probenvorbereitung weiter optimiert, um eine effektive Aufbereitung von klinischen Proben im Zusammenhang mit MSS und auch von Primärzelllinien sowie Säugerzellgeweben für anschließende, quantitative LC-MS Analytik zu ermöglichen.

Durch Nutzung dieser erstellten Arbeitsanweisungen, konnten tausende von Proteinen in MSS-Patienten und *woozy* relativ quantifiziert werden. Unter diesen Proteinen variierte die Menge von differenziell regulierten Proteinen abhängig von Zell- und Gewebstyp und vom klinischen Zustand, insbesondere durch deren Abhängigkeit von SIL1 oder Unabhängigkeit davon. Dennoch zeigte sich bei Abwesenheit von funktionellem SIL1 eine Störung der zellulären Aktivität in beiden Zuständen (SIL1 abhängig oder unabhängig). Dies deutet auf eine wichtige Funktion von SIL1 in der BiP-vermittelten Proteinfaltung hin. In MSS SIL1 unabhängigem Gewebe wurden die Prozesse, welche womöglich die Reduktion des ER-Stresszustandes herbeiführen und durch SIL1-Verlust induziert ist, identifiziert. Anschließend wurde eine Proteomanalyse von SIL1 *knock-down* HEK293 Zelllinien durchgeführt, um die Pathophysiologie im Falle von SIL1-Verlust im Detail zu untersuchen. Weiterhin wurde eine *targeted-MS* basierte Strategie entwickelt, um Proteine zu studieren, welche involviert sind in Signalwegen der Antwort auf ungefaltete Proteine und im Fall von ER-Stresssituation hervorgerufen wird. Letztlich wurde eine proteomische Analyse einer humanen RCMH Zelllinie durchgeführt. Diese Zelllinie kann womöglich als ein *in vitro* Modell zur Untersuchung von muskulären und neuromuskulären Fehlsteuerungen dienen.

Table of contents

Table of contents

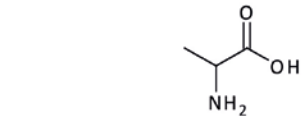
Abbreviations

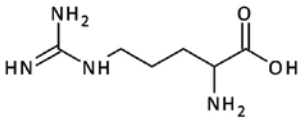
2-DE	two-dimensional gel electrophoresis
Å	angstrom
ACN	acetonitrile
ANOVA	analysis of variance
BCA	bicinchoninic acid
CID	collision-induced dissociation
Da	Dalton
DDA	data dependent acquisition
DTT	dithiothreitol
EBV	Epstein-Barr virus
e.g.	for example
EtOH	ethanol
FA	formic acid
FDR	false discovery rate
HCD	higher energy collisional dissociation
HEK293	human embryonic kidney cell line
HPLC	high performance liquid chromatography
HR/AM	high-resolution accurate mass
IAA	iodoacetamide
i.e.	that is
IPG	immobilized pH-gradient
IEF	isoelectric focusing
iTRAQ	isobaric tags for relative and absolute quantification
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSS	Marinesco-Sjögren Syndrome
MW	molecular weight
m/z	mass-to-charge
PRM	parallel reaction monitoring
PSM	peptide-spectrum match
ppm	parts per million
rcf	relative centrifugal force
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SPEC	solid phase extraction cartridge
SRM	selected reaction monitoring
TFA	trifluoroacetic acid
TIC	total ion chromatogram
u	atomic mass unit

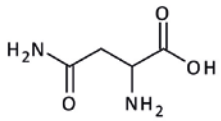
Note: Throughout the text, the symbols for human genes are italicized characters that are all in upper-case e.g. *SIL1* and the respective proteins are non-italicized characters e.g. SIL1. In case of mouse, only the first letter is in upper-case e.g. *Sil1* and Sil1 - gene and protein symbols, respectively.

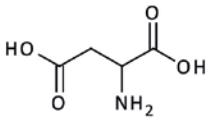
Chemical structures of amino acids

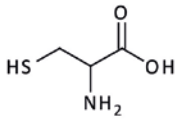
Name, three and single letter codes; monoisotopic residue mass

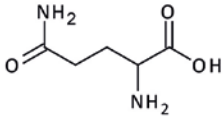

Alanine, Ala, A;
71.03711

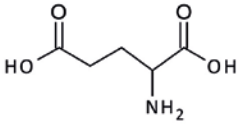

Arginine, Arg, R;
156.10111

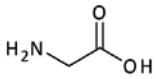

Asparagine, Asn, N;
114.04293

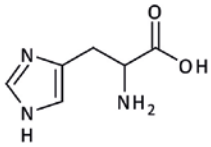

Aspartic acid, Asp, D;
115.02694

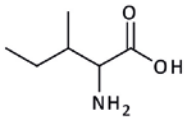

Cysteine, Cys, C;
103.00919

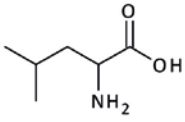

Glutamine, Gln, Q;
128.05858

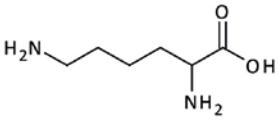

Glutamic acid, Glu, E;
129.04259

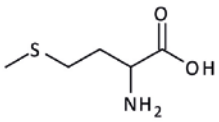

Glycine, Gly, G;
57.02146

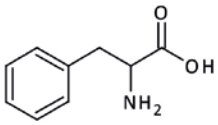

Histidine, His, H;
137.05891

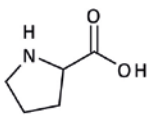

Isoleucine, Ile, I;
113.08406

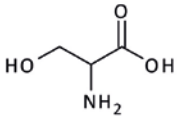

Leucine, Leu, L;
113.08406

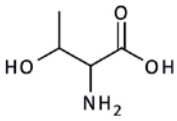

Lyine, Lys, K;
128.09496

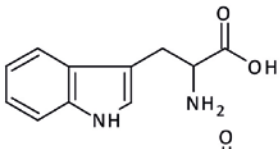

Methionine, Met, M;
131.04049

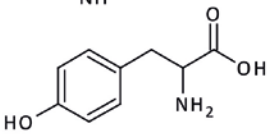

Phenylalanine, Phe, F;
147.06841

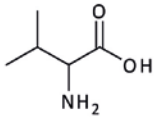

Proline, Pro, P;
97.05276


Serine, Ser, S;
87.03203


Threonine, Thr, T;
101.04768


Tryptophan, Trp, W;
186.07931


Tyrosine, Tyr, T;
163.06333


Valine, Val, V;
99.06841

Notes:

- Trypsin cleaves at the carboxyl (C)-terminus of K and R unless followed by P.
- iTRAQ labels covalently react with the primary amines on N-terminus of protein/peptide and K side chains.

List of publications

Publications related to this work

Kollipara, L.; Buchkremer, S.; González Coraspe, JA.; Senderek, J.; Weis, J.; Zahedi, RP.; Roos, A., In-depth phenotyping of lymphoblastoid cells suggests selective cellular vulnerability in Marinesco-Sjögren syndrome. *Oncotarget* **2016**, (under revision).

Roos, A.; **Kollipara, L.**; Buchkremer, S.; Labisch, T.; Brauers, E.; Gatz, C.; Lentz, C.; Gerardo-Nava, J.; Weis, J.; Zahedi, R. P., Cellular Signature of SIL1 Depletion: Disease Pathogenesis due to Alterations in Protein Composition Beyond the ER Machinery. *Mol Neurobiol* **2016,** *53* (8), 5527-41.

Kollipara, L.; Buchkremer, S.; Weis, J.; Brauers, E.; Hoss, M.; Rütten, S.; Caviedes, P.; Zahedi, R. P.; Roos, A., Proteome Profiling and Ultrastructural Characterization of the Human RCMH Cell Line: Myoblastic Properties and Suitability for Myopathological Studies. *Journal of Proteome Research* **2016,** *15* (3), 945-955.

Roos, A.; Buchkremer, S.; **Kollipara, L.**; Labisch, T.; Gatz, C.; Zitzelsberger, M.; Brauers, E.; Nolte, K.; Schroder, J. M.; Kirschner, J.; Jesse, C. M.; Goebel, H. H.; Goswami, A.; Zimmermann, R.; Zahedi, R. P.; Senderek, J.; Weis, J., Myopathy in Marinesco-Sjogren syndrome links endoplasmic reticulum chaperone dysfunction to nuclear envelope pathology. *Acta neuropathologica* **2014,** *127* (5), 761-77.

Kollipara, L.; Zahedi, R. P., Protein carbamylation: in vivo modification or in vitro artefact? *Proteomics* **2013,** *13* (6), 941-4.

Publications not related to this work

Psatha, K.; **Kollipara, L.**; Voutyraki, C.; Divanach, P.; Sickmann, A.; Rassidakis, G. Z.; Drakos, E.; Aivaliotis, M., Deciphering lymphoma pathogenesis via state-of-the-art mass spectrometry-based quantitative proteomics. *Journal of Chromatography B*. doi: 10.1016/j.jchromb.2016.11.005. [Epub ahead of print].

Reifschneider, O.; Marx, C.; Jacobs, J.; **Kollipara, L.**; Sickmann, A.; Wolters, D.; Kuck, U., An ribonucleoprotein supercomplex involved in trans-splicing of organelle group II introns. *J Biol Chem* **2016**.

Brauers, E.; Roos, A.; **Kollipara, L.**; Zahedi, R. P.; Beckmann, A.; Mohanadas, N.; Bauer, H.; Häusler, M.; Thoma, S.; Kress, W.; Senderek, J.; Weis, J., The Caveolin-3 G56S sequence variant of unknown significance: Muscle biopsy findings and functional cell biological analysis. *PROTEOMICS – Clinical Applications*. doi: 10.1002/prca.201600007. [Epub ahead of print].

Ciregia, F.; **Kollipara, L.**; Giusti, L.; Zahedi, R. P.; Giacomelli, C.; Mazzoni, M. R.; Giannaccini, G.; Scarpellini, P.; Urbani, A.; Sickmann, A.; Lucacchini, A.; Bazzichi, L., Bottom-up proteomics suggests an association between differential expression of mitochondrial proteins and chronic fatigue syndrome. *Transl Psychiatry* **2016,** *6* (9), e904.

Roos, A.; Buchkremer, S.; Coraspe, J. G.; Weis, J.; **Kollipara, L.**; Zahedi, R., Functions of the SIL1-BiP chaperone system in maintaining muscle fiber integrity. *Neuromuscular Disorders* 26, S211. http://dx.doi.org/10.1016/j.nmd.2016.06.454

Engel, J.; Becker, C.; Lategahn, J.; Keul, M.; Ketzer, J.; Muhlenberg, T.; **Kollipara, L.**; Schultz-Fademrecht, C.; Zahedi, R. P.; Bauer, S.; Rauh, D., Insight into the Inhibition of Drug-Resistant Mutants of the Receptor Tyrosine Kinase EGFR. *Angewandte Chemie* **2016,** *55* (36), 10909-12.

Fernando, R. N.; Chaudhari, U.; Escher, S. E.; Hengstler, J. G.; Hescheler, J.; Jennings, P.; Keun, H. C.; Kleinjans, J. C.; Kolde, R.; **Kollipara, L.**; Kopp-Schneider, A.; Limonciel, A.; Nemade, H.; Nguemo, F.; Peterson, H.; Prieto, P.; Rodrigues, R. M.; Sachinidis, A.; Schafer, C.; Sickmann, A.; Spitkovsky, D.; Stober, R.; van Breda, S. G.; van de Water, B.; Vivier, M.; Zahedi, R. P.; Vinken, M.; Rogiers, V., "Watching the Detectives" report of the general assembly of the EU project DETECTIVE Brussels, 24-25 November 2015. *Archives of toxicology* **2016,** *90* (6), 1529-39.

Solari, F. A.; **Kollipara, L.**; Sickmann, A.; Zahedi, R. P., Two Birds with One Stone: Parallel Quantification of Proteome and Phosphoproteome Using iTRAQ. *Methods in molecular biology (Clifton, N.J.)* **2016,** *1394*, 25-41.

Lefebvre-Legendre, L.; Reifschneider, O.; **Kollipara, L.**; Sickmann, A.; Wolters, D.; Kuck, U.; Goldschmidt-Clermont, M., A pioneer protein is part of a large complex involved in trans-splicing of a group II intron in the chloroplast of Chlamydomonas reinhardtii. *The Plant journal : for cell and molecular biology* **2016,** *85* (1), 57-69.

Limonciel, A.; **Kollipara, L.**; Zahedi, R.; Sickmann, A.; Jennings, P., Identification of novel mechanistic biomarkers of nephrotoxicity by proteomics and phosphoproteomics P10-015, *Toxicology Letters* **2015,** *238* (2 Supplement), S232. DOI: 10.1016/j.toxlet.2015.08.685

Venne, A. S.; **Kollipara, L.**; Zahedi, R. P., The next level of complexity: crosstalk of posttranslational modifications. *Proteomics* **2014,** *14* (4-5), 513-24.

Linxweiler, J.; **Kollipara, L.**; Zahedi, R. P.; Lampel, P.; Zimmermann, R.; Greiner, M., Proteomic insights into non-small cell lung cancer: New ideas for cancer diagnosis and therapy from a functional viewpoint. *EuPA Open Proteomics* **2014,** *4*, 25-39.

Sibilski, C.; Mueller, T.; **Kollipara, L.**; Zahedi, R. P.; Rapp, U. R.; Rudel, T.; Baljuls, A., Tyr728 in the kinase domain of the murine kinase suppressor of RAS 1 regulates binding and activation of the mitogen-activated protein kinase kinase. *J Biol Chem* **2013,** *288* (49), 35237-52.

List of poster presentations

Conference: 14th Human Proteome Organization (HUPO) World Congress, September-2015, Vancouver, Canada.
Title: "Quantitative proteomics reveals new insights into chaperone malfunction linked neurodegeneration"
Authors: Laxmikanth Kollipara*, Stephan Buchkremer, Joachim Weis, Andreas Roos, René P Zahedi

Conference: 12th Human Proteome Organization (HUPO) World Congress, September-2013, Yokohama, Japan.
Title: "Quantitative proteomics to study neuromuscular disorders"
Authors: Laxmikanth Kollipara*, Stephan Buchkremer, Andreas Roos, René P Zahedi

Conference: 61st Conference on Mass Spectrometry and Allied Topics (ASMS), June-2013, Minnesota, USA.
Title: "Extent of urea-induced protein carbamylation during sample preparation"
Authors: Laxmikanth Kollipara, René P Zahedi*

Conference: Proteomic Forum, March-2013, Berlin, Germany.
Title: "Quantitative proteomics to study neuromuscular disorders"
Authors: Laxmikanth Kollipara*, Stephan Buchkremer, Andreas Roos, René P Zahedi

Conference: The Society for Biochemistry and Molecular Biology (GBM), September-2011, Frankfurt, Germany.
Title: "Understanding Marinesco-Sjögren syndrome (MSS) using quantitative proteomics"
Authors: Laxmikanth Kollipara*, Andreas Roos, René P Zahedi

Conference: Proteomic Forum, March-2011, Berlin, Germany.
Title: "Understanding Marinesco-Sjögren syndrome (MSS) using quantitative proteomics"
Authors: Laxmikanth Kollipara*, Andreas Roos, René P Zahedi

Note: Underlined name* = presenting author.

1 Introduction

1.1 Endoplasmic reticulum and protein folding process

The endoplasmic reticulum (ER) also known as the sarcoplasmic reticulum (in skeletal muscle fibers) is the largest subcellular organelle in eukaryotic cells. Being a single membrane system, the ER is located closely to the nucleus and connected to it by the outer nuclear membrane of the nuclear envelope. Based on morphology and function, there are two types of ER i.e. smooth ER (sER) and rough ER (rER). The sER has a smooth outer surface and it is mostly involved in the synthesis and metabolism of phospholipids and steroids. In contrast, the outer membrane of the rER is constellated with the ribosomes, which harbor translational elongation factors that are regarded as the workhorses of protein biosynthesis [1]. Evidently, the rER (hereafter referred as only ER) is the site of production, folding and translocation of nearly one-third of all proteins in an eukaryotic cell [2]. It is estimated that the protein concentration in the ER lumen can reach up to 100 mg/mL [3]. Most importantly, the maturation of newly synthesized secretory and membrane proteins starts in the ER lumen. Properly folded proteins travel from the ER through the ER-Golgi-intermediate compartment (ERGIC) and the Golgi apparatus to their final destinations (e.g. cell surface or the extracellular space) [4]. The entry of incipient proteins into the ER for their maturation can be either by post- or co-translational transport [5]. During the co-translational protein translocation (Fig. 1.1), the nascent polypeptide chains are imported into the ER through the signal-gated SEC61 pore complex [6]. The signal peptide sequences are then cleaved off by the signal peptidase (Spase) and the residual polypeptide sequences are subjected to the protein-folding machinery, which mainly comprises ER-resident molecular chaperones and other folding enzymes. Chaperones are a group of proteins that assist folding and interactions between the polypeptide chains and thus gives rise to functional proteins [7]. The newly synthesized proteins are further stabilized by covalent post-translational modifications (PTMs) such as N-glycosylation and formation of the disulfide bonds, which are catalyzed by oligosaccharyltransferase and protein disulfide isomerases (PDI), respectively. Moreover, the ER compartment is a store house of cellular calcium (Ca^{2+}) and thus the ER plays a major role in maintaining Ca^{2+} dependent homeostasis [8]. Notably, the estimated available

(unbound) Ca^{2+} concentration in the ER can reach up to 1 mM [9]. Due to the presence of such diverse group of biomolecules including a large dynamic range of unfolded/folded proteins in a high oxidizing environment (crucial for the formation of disulfide bonds mediated by PDI) and high Ca^{2+} concentration; a strict ER-quality control (ERQC) mechanism has evolved and conserved in eukaryotes that determines the fate of newly synthesized polypeptide chains [10]. In summary, different components of the ER-protein folding machinery orchestrate folding/unfolding of the nascent polypeptides and thus prevent their aggregation and assembled macromolecular subunits [7] (Table 1.1).

Table 1.1: The main components (families) of the ER protein folding machinery [11].

Family	Prominent examples	Gene	Function(s)
Chaperones			
Heat-shock protein (HSP) 70	BiP	*HSPA5*	ERQC
	GRP170	*HYOU1*	Cytoprotection, Nucleotide exchange factor for BiP
HSP40	ERdj1-5	*ERDJ5*	Co-chaperone regulating ATPase activity of BiP
HSP90	Endoplasmin	*HSP90B1*	Processing and transport of secreted proteins
Lectins	Calnexin	*CANX*	Glycoprotein quality control
	Calreticulin	*CALR*	Glycoprotein quality control
Co-chaperones			
GrpE like	SIL1	*SIL1*	Nucleotide exchange factor for BiP
Peptidyl-prolyl isomerases (PPIs)			
Cyclophilins	Cyclophilin B	*PPIB*	cis-trans isomerization of proline
FK506-binding proteins	FKBP1A	*FKBP1A*	cis-trans isomerization of proline
Protein disulfide-isomerases	Protein disulfide-isomerase	*P4HB*	Formation / breakage of disulfide bonds

1.1.1 SIL1-BiP chaperone system

Among the multitude of proteins involved in the ERQC, the lumenal chaperone BiP - a member of the HSP70 family, plays a key role [12]. BiP, also called glucose related protein 78 (GRP78), is one of the most abundant ER-resident chaperones that is involved in various important regulatory processes. These include (i) folding/unfolding of nascent polypeptide chains and mature proteins, (ii) maintaining the ER-Ca^{2+} homeostasis, (iii) protein translocation across the ER-membrane, (iv) involvement in the ER-associated degradation (ERAD) and initiation of the unfolded protein response (UPR) pathway [13]. Like other HSP70 chaperones, BiP has an N-terminal nucleotide binding domain, which interacts with ATP/ADP and a C-terminal substrate binding domain that interacts with un-/misfolded proteins, respectively. Usually, the substrates bound by BiP are shielded from non-specific interactions with other unfolded proteins by their aggregation prone exposed hydrophobic regions, which are typically buried in the native state,

and act as BiP binding targets. In the ER lumen, BiP shuttles between ATP and ADP-bound states. The ATP-bound BiP has low-affinity but faster exchange rate of substrates, whereas the ADP-bound state has high-affinity and slow exchange rate of substrates. This BiP-ATP/-ADP interchange process is known as the ATPase cycle, which is controlled by the ER-resident co-chaperones (HSP40 proteins) and nucleotide exchange/releasing factors (NEFs) [14]. On one hand, HSP40s function in promoting the hydrolysis of ATP, which causes a conformational change that induces tight binding of BiP to substrates. The NEFs, on the other hand, catalyze the removal of ADP that allows BiP to release the substrates, which subsequently proceed towards the ERGIC and the Golgi apparatus to attain their final native states.

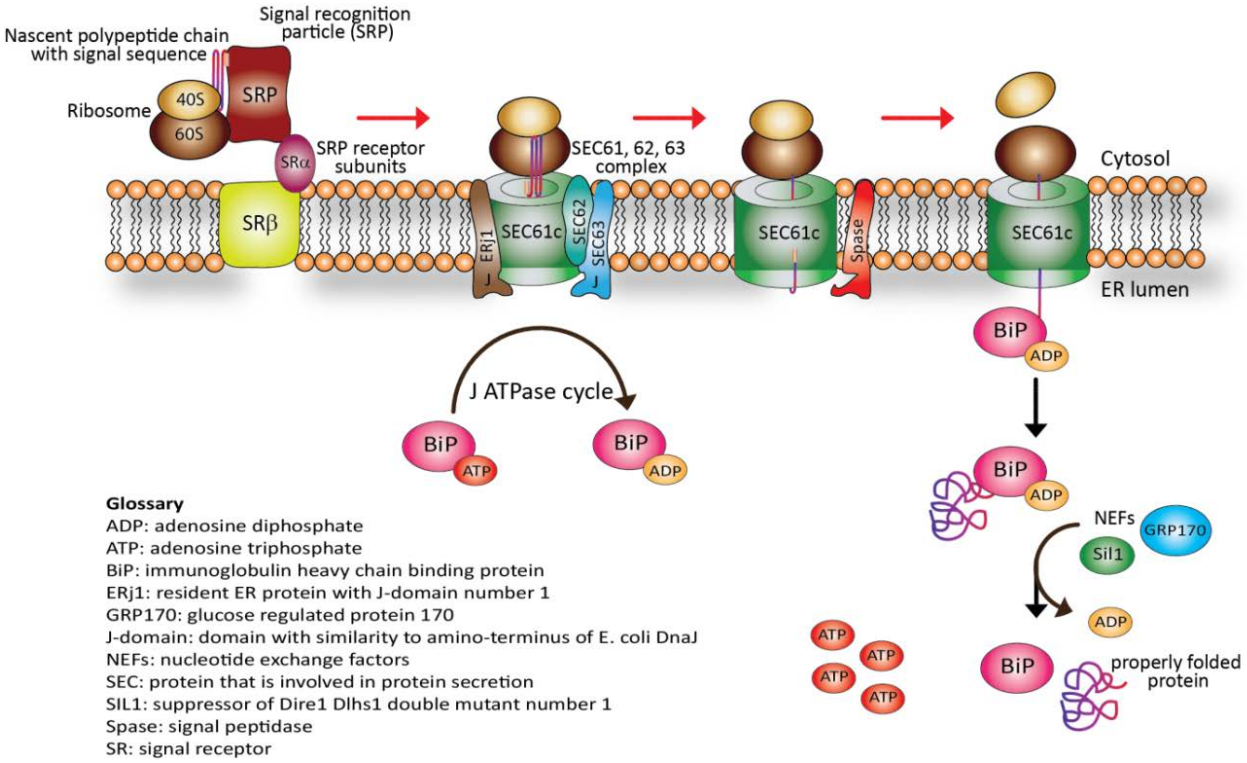


Figure 1.1: The co-translational protein transport pathway targets nascent polypeptides to the ER membrane via the signal recognition particle (SRP). In the ER lumen, BiP shuttles between ATP and ADP-bound states. This BiP ATP-ADP interchange process is known as the J ATPase cycle, which is regulated by the ER-resident co-chaperones and NEFs i.e. SIL1 and GRP170. Both NEFs displace ADP and subsequently release of the substrate (properly folded protein) from BiP to complete the protein folding process. The unbound BiP binds with the free ATP molecules in the ER lumen and the complete process is repeated. Figure adapted from [5].

There are two known NEFs for BiP i.e. (i) N-linked glycoprotein SIL1 (suppressor of Δire1 Δlhs1 double mutant number 1) and (ii) GRP170 (glucose related protein 170) also known as hypoxia up-regulated 1 [15]. Additionally, GRP170 has a cytoprotective function (under hypoxic conditions)

and can act as a chaperone [10], whereas no such chaperone activity has been reported for SIL1 so far. Notably, the analysis of the canine pancreatic rough microsomes revealed molar ratios of: 2.358 for BiP, 0.283 for GRP170 and 0.002 for SIL1 with respect to the alpha-subunit of SEC61 complex [16].Both SIL1 and GRP170 interact with the nucleotide binding domain of the ADP-bound BiP-substrate complex and displace ADP to subsequently release substrates from BiP [10, 17]. However, SIL1 preferentially binds to BiP and catalyzes the release of ADP that eventually completes the protein folding process [18] (Fig. 1.1). Structural analysis studies that were performed on SIL1-BiP complexes in yeast (*Saccharomyces cerevisiae*) have shown that the SIL1 ATPase domain interacts directly with the nucleotide binding domain of BiP with an equilibrium dissociation constant (K_d) of 13 nM [19]. Therefore, the SIL1-BiP interaction complex plays an essential role in the protein folding process and in maintaining cellular homeostasis. Evidently, the absence of SIL1 can cause decreased activity or loss of functional BiP and consequently the newly synthesized proteins might not be translocated into the ER leading to their accumulation in the cytosol, whereas the proteins that have entered the ER lumen remain in a un-/misfolded state and tend to aggregate. Furthermore, some of the important secretory and membrane proteins might fail to reach their final destinations and can cause secondary loss-of-functions [20].

1.2 ER stress, homeostasis and apoptosis

Many physiological and pathophysiological conditions such as altered protein folding capacity, disturbed N-glycosylation or disulfide bond formation, hypoxia, bacterial/viral infections or alterations of the ER-Ca^{2+} homeostasis can promote ER stress [21]. This term describes an imbalance between the cellular demand for ER function/homeostasis and capacity [22]. Regardless of the above mentioned causes, the protein folding machinery is often compromised under ER stress situation resulting in an increasing amount of un-/misfolded proteins in the ER lumen that tend to form insoluble aggregates (e.g. inclusion bodies) [23]. The underlying mechanism of aggregate formation is the interaction between hydrophobic regions of the partially folded intermediates, especially β sheets [24] leading to a non-native protein conformation that becomes devoid of its molecular function [23].

To alleviate the ER stress, two main stress response pathways are triggered to restore normal physiological functions in the eukaryotic cells i.e. (i) adaptive and (ii) apoptotic mechanisms [25].

The adaptive (or protective) response attempts to establish cellular homeostasis by three ways, which are regulated by the unfolded protein response (UPR) [26]. There are three ER-resident transmembrane proteins (known as UPR stress sensors) that transduce the unfolded protein signal across the ER membrane (i) double-stranded RNA-activated protein kinase (PKR)-like endoplasmic reticulum kinase (PERK), (ii) activating transcription factor 6 (ATF6) and (iii) inositol-requiring protein 1α (IRE1α), respectively [27] (Fig. 1.2).

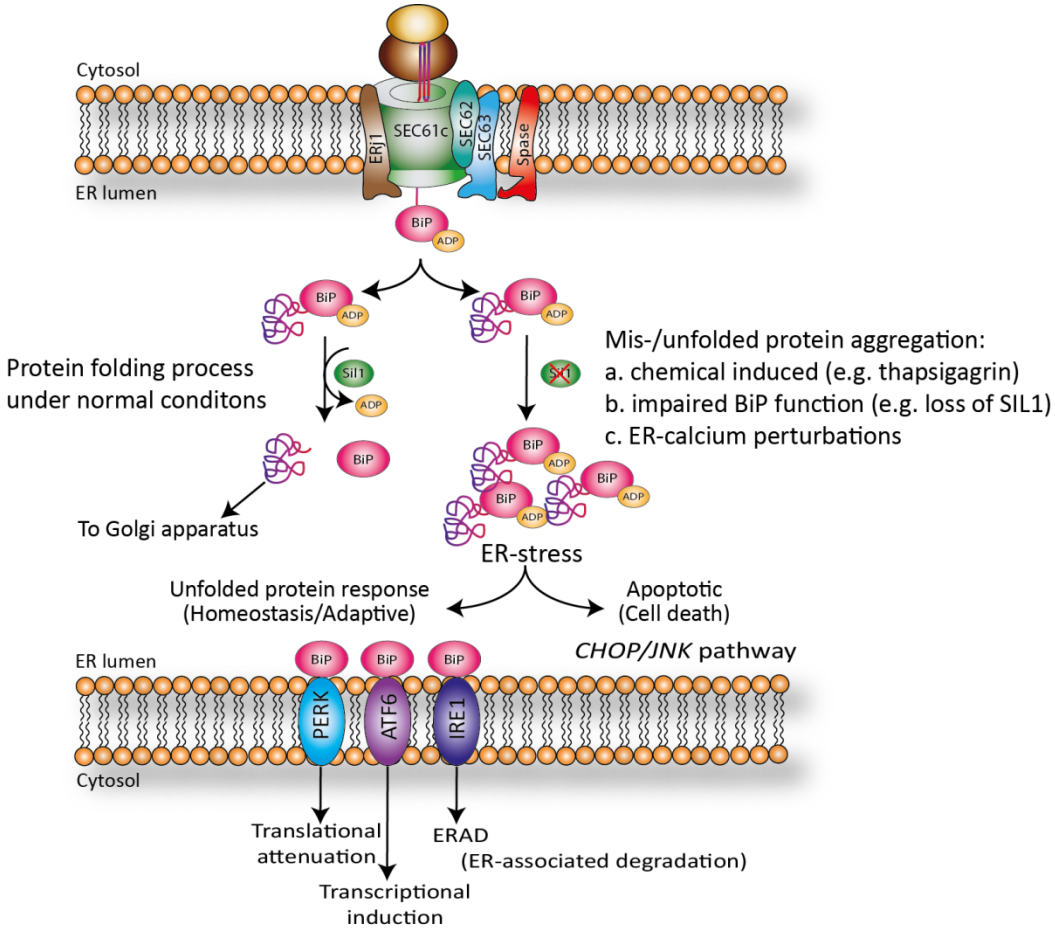


Figure 1.2: ER stress and activation of the UPR pathway. Under normal conditions, the ER folding machinery ensures proper maturation of proteins, which are further stabilized and targeted to their respective destinations by the secretory pathway. However, under the ER-stress conditions, aggregation of misfolded proteins occurs and the cell initially tries to restore homeostasis by triggering the UPR pathway (adaptive phase). The transmembrane BiP-bound UPR transducers (PERK, ATF6 and IRE1) modulate cellular homeostasis by downstream signaling pathways; however, under persistent stress conditions, the cell undergoes the apoptotic (or programmed cell death) phase [28].

These three signaling sensors of the UPR pathway operate in parallel and they are supposed to sense the ER-stress through BiP binding/release by their respective lumenal domains [25]. The first protective mechanism is mediated by PERK, which is activated by dimerization and

phosphorylation. Once activated, PERK phosphorylates initiation factor eIF2, resulting in translation attenuation to prevent further synthesis of nascent proteins. The second mechanism involves ATF6, which is sequentially cleaved into two subunits: ATF6α and ATF6β by two distinct proteases i.e. site-1 protease and site-2 protease, respectively that are located in the Golgi complex. Both subunits then translocate into the nucleus and positively promote transcription of the ER stress target genes including the ER-resident molecular chaperones and the folding enzymes [29]. The last adaptive mechanism is mediated by IRE1, which is activated by oligomerization and autophosphorylation. Upon activation, it regulates (i) chaperone induction, (ii) the ERAD and (iii) enlargement of the ER surface area to accommodate the bulk of unfolded proteins [11b]. The ERAD machinery specializes in recognizing the misfolded proteins and (retro)translocates them to the cytoplasm for their degradation by the ubiquitin-proteasome pathway [30]. This pathway involves two steps: First, proteins destined for degradation are subjected to ubiquitination, a PTM in which a single or multiple ubiquitin protein molecules are attached to the Lys residues of the substrate proteins. Second, ubiquitin-tagged proteins are proteolytically degraded in the 26S proteasome complex [31] or occasionally, by the lysosomes/vacuole [32]. However, under prolonged and persistent ER-stress conditions; the adaptive response pathway ceases and usually gives way to the apoptotic pathway, which is mediated by C/EBP homologous protein (CHOP) or c-Jun N-Terminal kinase (JNK) leading to programmed cell death [33].

1.3 Protein aggregation related mammalian disorders

Accumulation and aggregation of un-/misfolded proteins is cytotoxic and causes tissue and organ damage that can lead to progressive development of various neurodegenerative [23, 34] neuromuscular and muscle disorders in mammals [35]. A collective term known as "conformational" or "folding" diseases describes the pathological conditions caused by abnormally folded cellular proteins and their aggregation [34b, 36]. Some of the well-known examples include Alzheimer's disease, Huntington's disease and Parkinson's disease [34a]. Over the years, clinical evidences accumulated from many patients' biopsy who suffered from these diseases have revealed the presence of intra-/extracellular aggregates of un-/misfolded proteins or mutated gene products (non-functional proteins) [37]. Besides, studies conducted on human

and mouse models of neurodegeneration and neuromuscular degeneration showed altered levels of chaperones, UPR sensors and apoptotic mediators, which indicate the ER stress related activation of the ERAD and UPR pathways and disturbed protein clearance processes. Notably, these mechanisms were found to be critical for the survival of cells especially in neurodegenerative disorders in man and mouse models [37-38].

1.3.1 Marinesco-Sjögren Syndrome

Marinesco-Sjögren Syndrome (MSS, OMIM: 248800) is a progressive multi-systemic neuromuscular disorder first described in 1931 by Gheorge Marinesco [39]. It is a rare genetic autosomal recessive condition with phenotypic variability [40] and so far 200 cases of MSS were reported according to the Orphanet Reports Series 2015 (http://www.orpha.net/consor/cgi-bin/Education_Home.php). In the past decade, genetic studies have identified mutations in *SIL1* (chromosome *5q31*) as the main cause of MSS in nearly 60% of the examined cases [41]. The clinical symptoms of MSS patients are: mental impairment, bilateral congenital or infantile cataracts, marked vacuolar myopathy and cerebellar ataxia [42]. However, ataxia, cataracts and myopathy represent the characteristic "clinical triad" or "classical MSS-phenotype" [41a] (Fig. 1.3 A-C).

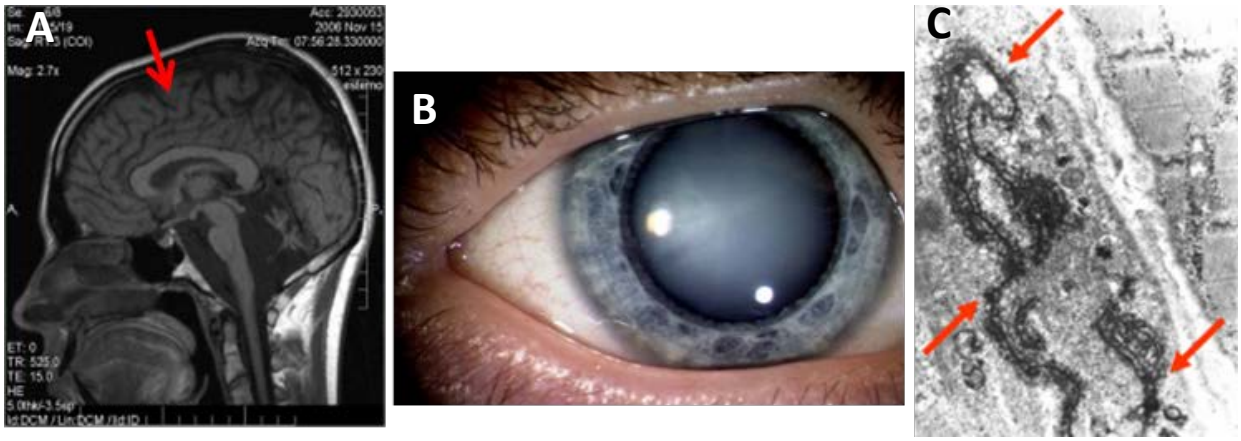


Figure 1.3: Clinical and electron microscopy findings in MSS patients - the characteristic "clinical triad". **(A)** Marked cerebellar atrophy in patient MSS33 at age 24 years (T1-weighted magnetic resonance imaging, MRI). **(B)** Cataract in a 7-year old patient [41a]. **(C)** Dense membranous structures surrounding the myonucleus (red arrows; biopsy of a 15-year old patient) scale 1 µm, respectively [41a, 41b].

Mental retardation can be of varying degree or maybe absent [41a]. Additional clinical symptoms in some MSS cases are short stature, hypogonadism, scoliosis, nystagmus and strabism ("non-

classical MSS-phenotype") [41a]. However, a small proportion of cases with the "classical" and a high proportion of cases with "non-classical" MSS-phenotypes do not have *SIL1* mutations thereby indicating a heterogeneity of this phenotype. Notably, mutations within the functional candidate genes *HSPA5* (BiP) and *HYOU1* (GRP170) were already excluded by molecular genetic studies of *SIL1* mutation-negative MSS patients. Besides these phenotypic alterations, morphological studies on the muscle specimens of most MSS patients showed severe mitochondrial alterations and myonuclear irregularities. The presence of irregular electron-dense membranes surrounding a part of the diseased myonuclei is a striking feature of MSS-related myopathy [35, 43] (Fig. 1.3 C). Recently Krieger et al., [41a] confirmed the presence of 46 different types of mutations on *SIL1*, further consolidating the wide mutational spectrum. Depending on the type of mutation, the functional SIL1 can be (i) completely absent, (ii) truncated (loss of critical functional domains) or (iii) destabilized and thus leading to an impaired BiP-associated protein folding process [41a, 44]. Importantly, in man, mutations in *SIL1* show selective vulnerability affecting only certain tissues/organs, which include the brain, the eyes and skeletal muscles [45]. Notably, the muscular and central nervous vulnerability can be confirmed in mice [34c, 35] whereas the eyes do not seem to be affected by SIL1-deficiency [46]. The plausible speculation for this behavior could be that certain cell types and organs are fully dependent on the functional SIL1 in the BiP-mediated protein folding process [5, 18].

Proteomic analyses of the MSS patients and an index patient

Proteomic studies of two different cell types that were derived from clinically unaffected tissues of five different MSS families were performed in this work (Table 1.2). Genetic analysis of the fibroblasts cultures from the skin biopsies and lymphoblastoid cells taken from these MSS patients have revealed the presence of following *SIL1* mutations: exon skipping (MSS2), small in-frame exon deletions (MSS24), missense mutation (MSS32), frameshift mutation in the last exon (MSS64), and nonsense mutation (MSS33) [41a, 41b]. These mutations result in the expression of SIL1 protein, albeit in low levels and mostly in the non-functional form as the domains responsible for SIL1-BiP interaction are either absent in the expressed protein or the residual/truncated SIL1 exhibits weak binding with BiP [41a, 41b].

In addition to these five MSS cases, proteome analysis of the muscle lysate derived from an index patient (female, Caucasian) who presented phenotypic manifestations similar to MSS, but not related to *SIL1* mutations was also performed in this work. Apart from *SIL1*, mutations of all limb girdle muscular dystrophy related genes were excluded as to be the disease causative factor of the patients' phenotype. Moreover, these symptoms were also identical to a distal myopathy 1 disorder (OMIM: 160500), which is caused due the mutations in *MYH7* that encodes for the cardiac β-myosin heavy chain or myosin-7 protein mostly found in the cardiac (heart) and skeletal muscles. Genetic analysis of the index patient DNA revealed a single nucleotide polymorphism in *MYH7*. However, this variant is considered as clinically non-relevant (http://www.ncbi.nlm.nih.gov/clinvar/variation/43056/#supporting-observations) and does not correlate with the severe phenotypic abnormalities showed by the index patient. In spite of the non-similar genetic defect, overlapping clinical symptoms with MSS, such as prominent myopathic changes including the considerable presence of protein aggregates, led to perform a comparative proteomics investigation by making use of skeletal muscle protein extract to gain preliminary insights into the myopathic changes of the index patient with respect to healthy controls.

Table 1.2: Ethnic details and clinical features of MSS patients with *SIL1* mutations [41a, 41b]. **Table legend:** del = deletion; het = heterozygous; hom = homozygous; mat = maternal; pat = paternal; + = present; - = absent; n.a = not available; P = proximal limbs; D = distal limbs.

Family	**MSS2**	**MSS24**	**MSS32**	**MSS33**	**MSS94**
Sex	male	female	male	male	male
Age	5 years	4 years	6 years	26 years	28 years
Ethnic origin	Turkey	Turkey	United States	Italy	Pakistan
SIL1 mutation(s)	645+1G-A, skipping, exon 6 (hom)	p.V231_l232del (hom)	p.G312R (mat); p.F345fs (pat)	p.S256fs (hom)	p.E101fs (hom)
Cataracts (age at diagnosis)	+ 4.5 years	-	+ 4 years	+ 3 years	+
Ataxia	+	+	+	+	+
Cerebellar atrophy	+	+	+	+	+
Muscle weakness	n. a.	+	-	P=D	-
Muscle atrophy	n. a.	+	P=D	P=D	-
Muscle biopsy findings	myopathic	myopathic	myopathic	myopathic	myopathic
Membranous structure associated with myonuclei	+	n. a.	n. a.	+	+
Cell type used for proteomics analysis	fibroblasts	lymphoblastoid cells			
Clinical state	unaffected				

Non-SIL1 related index patient: The clinical symptoms were: progressive microcephaly at 4 months; moderate gait ataxia with weak tendon reflexes and moderately elevated creatinine kinase level (365 U/L) at 2 y indicative of neuromuscular disorder; progressive generalized muscle weakness and myopathic facies between 3 to 5 years of age; from the age above 5, she was completely wheelchair-bound. The muscle biopsy for morphological and proteomics analyses was taken at 5 years of age.

1.3.2 Animal model of MSS

In 2005, Zhao and colleagues [34c] described a murine phenotype caused by mutant transcripts of *Sil1* (located on murine chromosome 18). These transcripts demonstrate splicing between exon 7 of *Sil1* and an ETn retrotransposon inserted between nucleotides 4,799 and 4,800. The chimeric transcript includes an in-frame stop codon after 96 nucleotides of ETn sequence. These homozygous mutant animals (commonly referred as *woozy* mice) suffer from cerebellar atrophy due to loss of the Purkinje cells from ages 3 to 4 months [34c, 35, 47] (Fig. 1.4 A, B) and develop advancing myopathy [35].

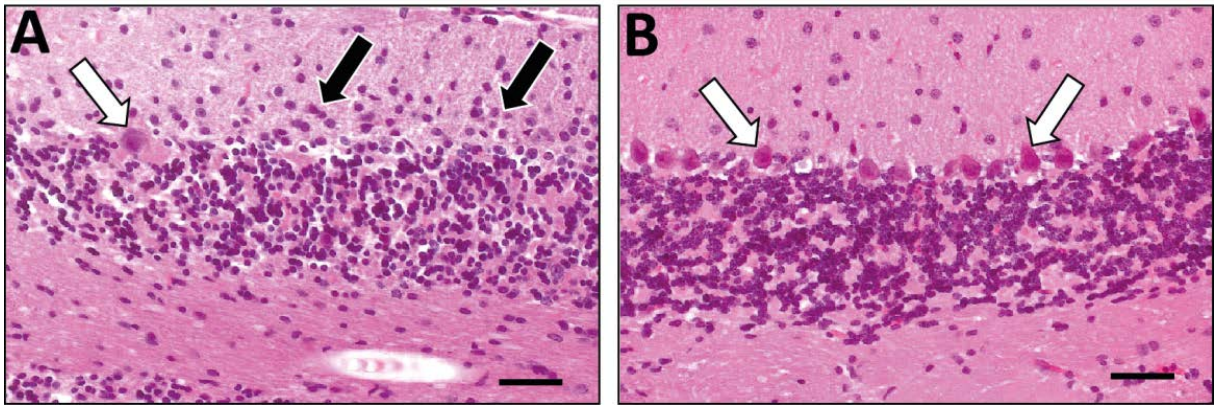


Figure 1.4: Immunohistochemistry (IHC) findings show **(A)** Purkinje cell loss and proliferation of glial cells in the cerebellum of Sil1-deficient *woozy* mouse (black arrows: proliferated Bergmann-glia; white arrow: a single survived Purkinje cell). **(B)** Regular density and arrangement of Purkinje cells in the cerebellum of wild type littermate [35] (white arrows). Scale 1 µm.

SIL1 is widely expressed in the brain, but its loss evidently affects the cerebellar Purkinje cells causing fatal consequences including autophagy and apoptosis [34c] leading to gait ataxia. Interestingly, it was shown that in *woozy* mice, concomitant overexpression of GRP170 rescues the Purkinje cell loss [47]. However, the exact role of the increased levels of GRP170 in Purkinje cells is still uncertain owing to the fact that GRP170 has a dual function i.e. chaperone activity as well as NEF for BiP [10]. Recently, Inaguma et al., [48] examined the pathophysiological significance of *SIL1* mutations in corticogenesis of MSS by conditional RNAi-based *Sil1*-silencing in mice. Depletion of murine SIL1 caused neuronal migration delay during corticogenesis, which could be rescued by concomitant expression RNAi-resistant Sil1, but not by three MSS-causing *Sil1* mutants. Moreover, SIL1-BiP interaction as well as BiP function was found to be crucial for neuronal migration *ex vivo*. Besides, time-lapse imaging revealed morphological disorganization

associated with abnormal migration of SIL1-deficient neurons in mice. Based on their findings, the authors concluded that abnormal neuronal migration and interhemispheric axon development may contribute to mental impairment in the pathophysiology MSS.

Therefore, due to the genetic background similarity and phenotypic resemblances, the *woozy* mouse serves as a suitable *in vivo* model to study the functional role of SIL1 in MSS [10].

1.4 Proteome analysis - in general

Unlike the genetic material (deoxyribonucleic acid or DNA), which is predominantly static and stable; the protein composition of a cell is highly dynamic and complex [49]. Slightly over two decades ago, Marc Wilkins coined the word "proteome" to describe the total protein population of a cell at a given point of time [50]. Recently, the human genome project provided a wealth of information about the human genetic code and revealed the number of protein-coding genes i.e. ~25,000 [51]. However, the exact number of genes is still under debate and it is believed to fluctuate in the future [52]. Interestingly, it has been suggested that these genes code for up to an estimated 1,000,000 distinct proteins. This disagreement between gene expression and the actual form of biologically active protein cannot be explained by transcriptomics, which is the study of the complete set of messenger ribonucleic acid or mRNA transcripts ("transcriptome") expressed by an organism's genome at a particular point in time. However, the large discrepancy between genes and proteins is attributed to factors such as alternative splicing and PTMs of proteins [53]. Therefore, it could be argued that studying proteins in living organisms provides more biological information with respect to gene expression because most of the metabolic and cellular processes are carried out by proteins. In contrast, the mRNA most often serves as an intermediate between genetic information and the functional proteins.

Following the footsteps of the human genome project, the human proteome project (HPP) was initiated in the year 2010, which aimed to identify all the proteins encoded by 20,300 genes. The key objectives of the HPP are (i) identification of protein isoforms as well as PTMs, (ii) quantification of differentially regulated proteins in control/disease states and (iii) annotation of the proteins to their biological function. Notably, one of the three main driving forces of the HPP apart from protein capture (antibody-based) profiling and protein information databases is mass spectrometry (MS) based proteomics [54]. Owing to the rapid technological developments in

instrumentation, the current MS-based proteomics workflows are capable of identifying several thousands of proteins in a span of few days [55] to just an hour [56] under ideal conditions. On the one hand, fast MS-based proteomics analysis is providing substantial amount of qualitative (identification) information using minute (µg) sample amounts and better analysis than two dimensional gel-based (i.e. 2-DE) approaches of complex samples such as cell lysates and PTMs (e.g. phosphorylation). On the other hand, quantitative proteomics techniques have gradually matured and their application to study human disorders has gained momentum [57] in the past 15 years. Moreover, cancer and clinical biomarker research areas have also been consistently utilizing quantitative proteomics approaches for understanding the molecular basis of disease mechanisms and to identify novel disease markers, respectively [58] (Fig. 1.5).

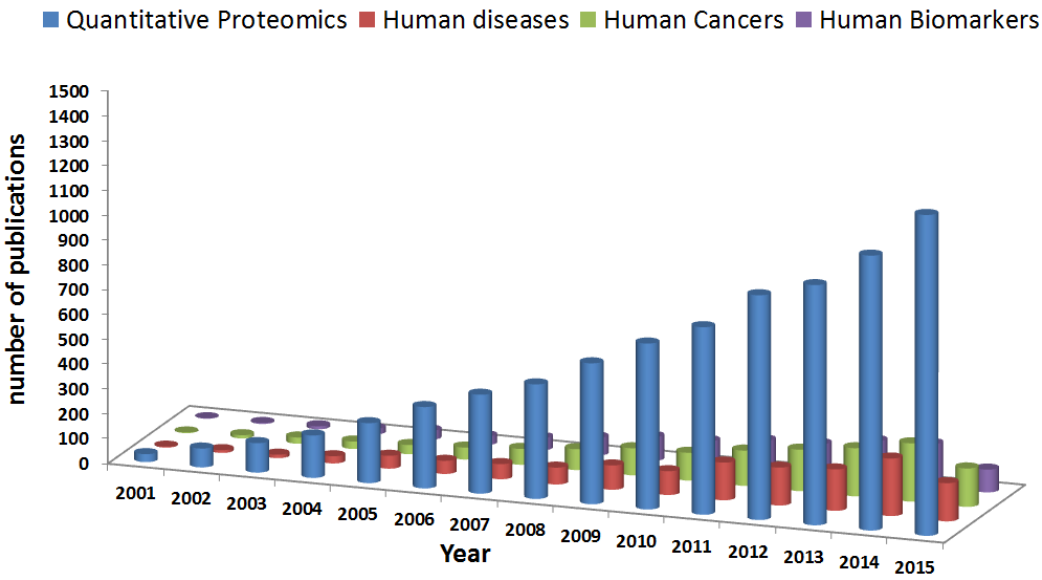


Figure 1.5: Number of accumulated research articles containing words "quantitative proteomics" only or in combination with human diseases/cancers/biomarkers in their abstracts. Overall, there has been a steady increase in all aspects since January-2001 till December-2015 (http://www.ncbi.nlm.nih.gov/pubmed).

1.4.1 Mass spectrometry - based proteomics

Characterization of proteins and peptides by mass spectrometry was boosted with the invention of two soft ionization procedures namely, matrix-assisted laser desorption ionization (MALDI) [59] and electrospray ionization (ESI) [60]. Both techniques can transform biomolecules (e.g. proteins and peptides) into gas phase ions in a non-destructive way to allow subsequent analysis by a mass spectrometer (MS). However, in recent years, ESI has become the method of choice in the

field of MS-based proteomics mainly because of its ease to couple high-performance liquid chromatography (HPLC) separations with a MS [61].

1.4.1.1 Electrospray ionization

The process of generating ions by ESI is carried out under atmospheric pressure and the molecules to be ionized are in a liquid state (Fig. 1.6).

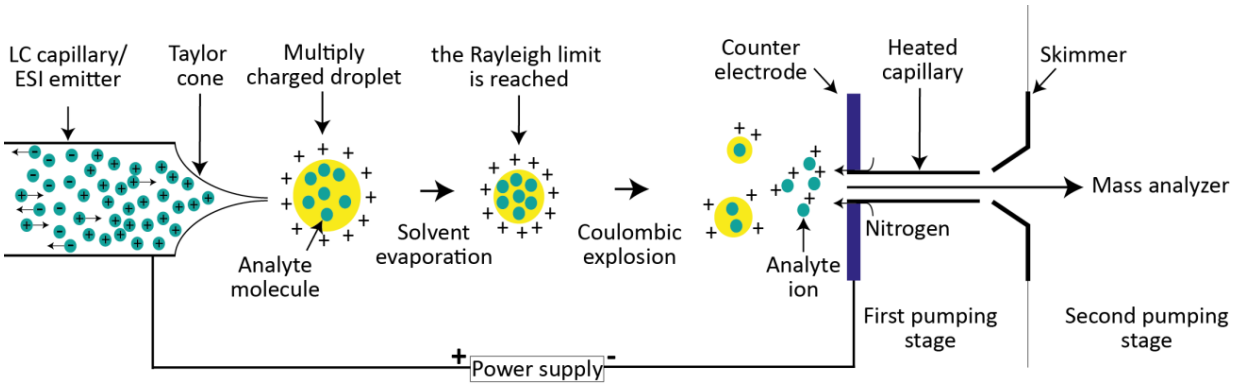


Figure 1.6: Schematic representation of ESI process. Gas-phase ions are produced by a repeated process of solvent evaporation and Coulombic explosions of multiply-charged droplets. The ions are subsequently transferred into MS which is usually operated under high vacuum (10^{-3} - 10^{-10} Torr) for two reasons: (i) to avoid collisions with neutral gas molecules and (ii) to increase the mean free path of ions. A feature of ESI is that highly multiply charged ions can be produced from these droplets, thereby allowing the analysis of biomolecules such as proteins and peptides by mass spectrometers with limited m/z range (e.g. 2,000). Figure adapted from: http://www.bris.ac.uk/nerclsmsf/techniques/hplcms.html.

The analyte solution is passed through a capillary (an ESI emitter), which is usually connected to an HPLC (LC-MS). A potential difference of usually ± 1.5 - 4.5 kV is applied between the emitter and the counter electrode. This strong electric field induces charge accumulation at the liquid surface located at the tip of the emitter (inner diameter ~10 µm). In positive ESI mode, the accumulated charges (positive) are attracted towards the counter electrode (negative) due to the electrostatic repulsion from the emitter (same polarity). When the voltage reaches a threshold value, known as the Taylor voltage, the liquid surface changes rapidly its shape to become a rounded cone - the "Taylor cone" and from the tip of this cone the liquid is sprayed (liquid filament). This process produces charged droplets and as the solvent evaporation occurs, the droplet shrinks until it reaches a point such that the surface tension can no longer sustain the charge repulsions ("Rayleigh limit"). At this point a "Coulombic explosion" of the multiply charged droplet occurs and a series of smaller (secondary) droplets are ejected. These droplets

again shrink due to evaporation and emit even smaller droplets. Furthermore, formation and desolvation of the droplets is aided by a heated capillary and in some cases by nebulizing gas (Nitrogen) flow at the MS inlet. This process of solvent evaporation and Coulombic explosion occurs repeatedly to generate smaller and secondary droplets until peptide ions are transferred to the gas phase - the mechanism of which is not completely revealed to date [62]. ESI has proven effective in producing gas phase ions of proteins and peptides. Notably, multiple charge states are often observed and the number of charges per biomolecule analyte typically depends on (i) the length of amino acid chain, (ii) the number of ionizable functional groups and (iii) the pH of HPLC solvents. Furthermore, a technical improvement of ESI known as the nano-electrospray ionization (nano-ESI) works with small analyte volume i.e. microliter (µL) range and at nanoliter/min flow rate has shown to increase the sensitivity of the analysis [63].

1.4.1.2 Reversed-phase chromatography

In so-called bottom-up proteomics, MS analysis is performed on the peptide level typically after proteolytic digestion of proteins using suitable proteases (e.g. trypsin). However, for highly complex proteomes e.g. human, a fractionation step prior to MS analysis is necessary. LC-MS mostly done with reversed-phase (RP) chromatography and MS - is a powerful combination to reduce the sample complexity of peptide mixtures [61] (Fig. 1.7).

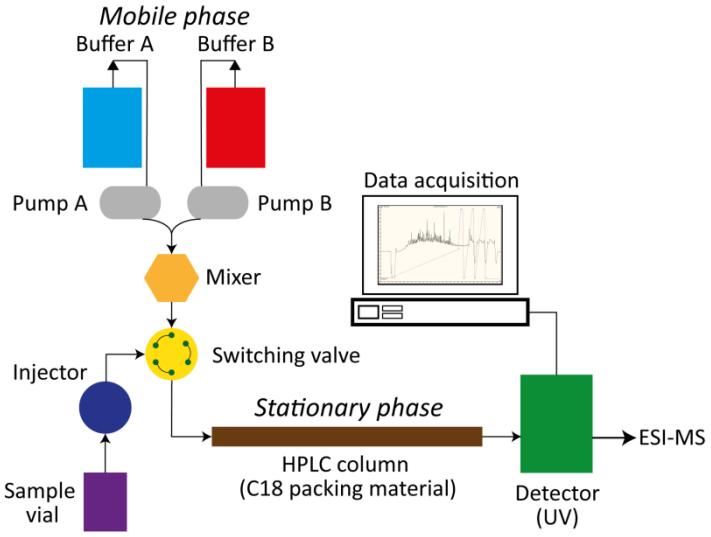


Figure 1.6: Block diagram of a HPLC system with a reversed-phase column (for peptide separations) and a UV detection system.

In RP, the peptide solutions are separated based on their varying degree of reversible hydrophobic interactions with a non-polar stationary phase in a polar mobile phase. The RP stationary phase column normally consists of uniform porous silica particles that are bonded with long chain alkyl groups (e.g. octadecyl/C18) whereas, the mobile phase include polar solvents such as water, methanol or acetonitrile. Peptide sequences that are mainly composed of hydrophilic amino acids preferably interact with the polar phase and hydrophobic amino acid containing peptides will adsorb strongly to the RP. The elution of peptides from the stationary phase is usually done by changing the composition of the mobile phase. Because of their diverse binding affinities, the preferred elution method of peptides from the RP stationary phase is performed by gradually increasing the organic concentration of the mobile phase over time (i.e. "gradient elution"). Often, the peptides eluting from the RP column pass through a detection system that monitors the separation usually a UV-wavelength detector, which measures the absorbance of the peptide bond at a wavelength of ~214 nm. In case of an LC-MS setup, the effluent from the column (through the UV detector) is coupled directly to a MS using an ESI interface.

1.4.1.3 Off-line fractionation strategies

In spite of its routine application, one-dimensional RP is still not sufficiently efficient to separate complex peptide mixtures. For this reason, multi-dimensional separation techniques that can exploit the physicochemical properties (e.g. charge, isoelectric point, hydrophobicity, molecular weight) of proteins/peptides have been developed [64]. Usually, the first dimension separation is carried out by gel-based or chromatography-based techniques followed by second dimension LC-MS analysis. In the present work, a modified version of isoelectric focusing known as the OFFGEL fractionator (Agilent) was used, which enables to recover the separated proteins/peptides fractions (24) in the liquid phase [65]. These individual fractions can be subsequently analyzed by LC-MS (Fig. 1.8 A).

Alternatively, two dimensional liquid chromatography (2D-LC) methods have been developed to improve the proteome coverage [66]. The effectiveness of any chromatographic method depends on its "separation power", which is described by the peak capacity i.e. the total number of peaks that can be fit into a chromatogram when every peak is separated from adjacent peaks. In order

to maximize the peak capacity, 2D-LC techniques often employ two chromatographic methods that have dissimilar (i.e. orthogonal) separation mechanisms or selectivities. For instance, the multi-dimensional protein identification technology (MudPIT) introduced by Yates and co-workers involves strong cation exchange (SCX) chromatography in the first dimension and RP in the second dimension prior to MS analysis [67]. Another 2D-LC strategy employs RP-HPLC separations in both dimensions, which are operated at two different pH values [68] (Fig. 1.8 B). First, the peptide mixtures are fractionated off-line by RP at basic pH (~10.0) and afterwards each fraction is subjected to LC-MS analysis, which is usually done at a low pH (~2.0). Although the chromatographic properties of peptides differ substantially between pH 2.0 and 10.0, this approach is considered as semi-orthogonal since it uses RP-RP chromatography in both dimensions. Nevertheless, Gilar and co-workers could demonstrate the dramatic change of charge distribution within the peptide chain due to the high-/low- pH values of RP-HPLC mobile phase [69]. Furthermore, concatenating several early, middle, and late fractions of high-pH RPC could minimize the overlap between the first and second dimensions and should improve overall proteome coverage. Recent studies have demonstrated that the orthogonality of concatenated high-/low- pH RP-RP fractionation is not only comparable to SCX-RP-HPLC, but also yielded slightly higher protein sequence coverages [70].

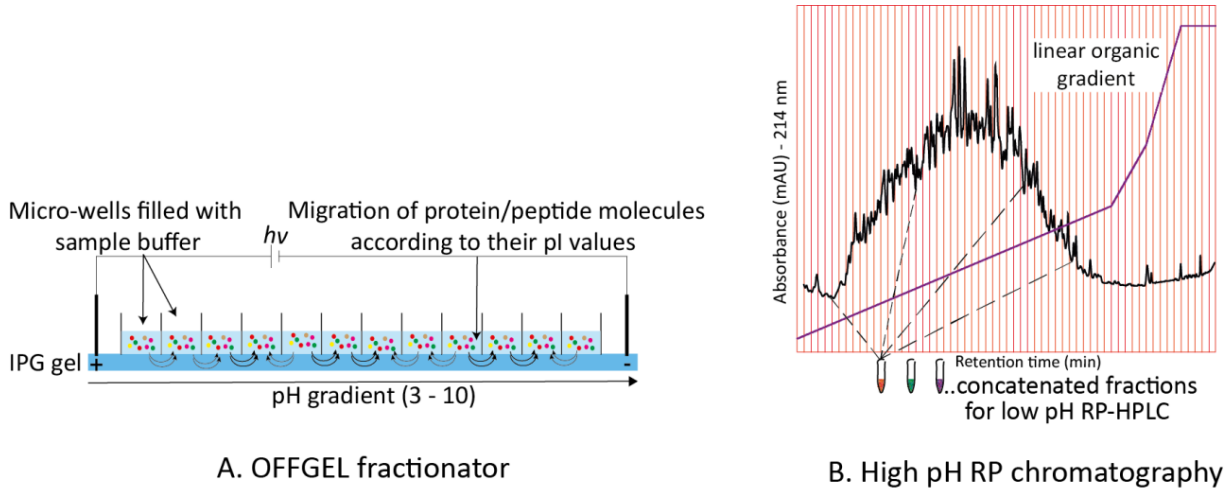


Figure 1.8: Examples of off-line fractionation strategies. **(A)** Protein or peptide separations are performed in a multi-compartment device called OFFGEL fractionator. In a pH gradient (3 - 10) the sample components migrate towards the anode or the cathode until they reach the pH values, where their net charge is zero: their isoelectric points (pI). The resulting fractions are in solution, which are recovered for subsequent LC-MS analysis. **(B)** UV chromatogram depicting the separation of a peptide mixture on a RP column using a linear organic gradient and operated at a high pH (6.0) HPLC. Each fraction is collected at regular retention time intervals, typically 60 s and to improve the separation orthogonality, several fractions from different elution time points are concatenated prior to LC-MS analysis.

1.4.1.4 Mass analyzers

MS usually consist of the following parts: (i) ion source (e.g. ESI) and optics, (ii) mass analyzer, (iii) detector and (iv) data processing electronics. Mass analyzers are an integral part of each instrument because they separate gas phase ions according to their mass-to-charge (m/z) ratios. Almost all mass analyzers use electric or electromagnetic fields to separate ions and the difference between the devices lies in the manner in which these fields are applied. Mass spectra are constructed by plotting ion intensity as a function of m/z ratio. A great variety of mass analyzers have been developed in the past century and each mass analyzer has unique performance characteristics. Some of these parameters have direct implications in the proteomics field [71] and they are briefly described below.

The **resolution (R)** of a mass analyzer is its ability to distinguish between two neighboring ions/peaks that differ only slightly in their mass (i.e. Δ m or Δ m/z). Mathematically, it is the inverse of resolving power given as m/Δ m. The most commonly used method to measure the resolution of a mass analyzer follows the full-width, half-maximum (FWHM) definition, which uses the width of a single peak at 50% of its height to determine Δ m (Fig. 1.9 A). In LC-MS based proteomics, high resolution mass analysis enables (i) to distinguish co-eluting or overlapping peptides, (ii) to accurately determine the charge states of the peptide ions and (iii) to determine their molecular masses (or m/z values) (Fig. 1.9 B).

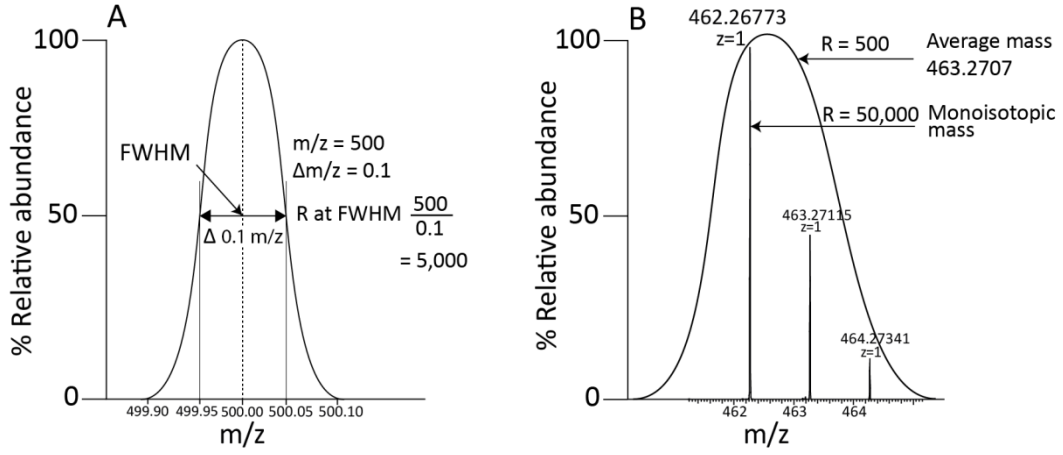


Figure 1.9: (A) FWHM method for determining resolution for a mass spectrometer measured at a given ion. **(B)** Comparison of two resolution values. High R not only enables to resolve the isotopic peaks but also for charge state determination and accurate mass determination.

High resolution MS usually refers to R values > 10,000 however, the resolution varies with the m/z value and it is instrument dependent. **Mass accuracy** is defined as the difference between the measured (experimental) mass and its calculated (theoretical) value, which is usually described in a relative manner e.g. part per million (ppm). Accurate mass measurement (1 - 2 ppm) allows a more confident identification of analytes. Although there is no direct correlation between mass accuracy and resolution, a correctly calibrated instrument with high resolution (e.g. R ≥ 60,000) can provide ≤ 1 ppm mass accuracy. **Sensitivity** and **dynamic range** are closely related terms associated with a MS, which give an indication of the maximum range of analyte concentrations that can be detected in a given sample. These are important parameters for proteomics analysis as biological samples have large differences in analyte concentrations. **Analysis speed** or **scan rate** of a mass analyzer refers to how fast it scans a mass spectrum. This is important for LC-MS applications where the chromatographic peak widths are typically between 5 - 10 s FWHM.

Commonly used instruments in proteomics are quadrupole (Fig. 1.10 A) and linear ion trap (Fig. 1.10 B) mass spectrometers that offer high sensitivity and fast scan rates depending on their mode of operation. The downside however, is both analyzers generate low resolution (R usually 2,000 and 4,000, respectively) and low mass accuracy (~100 ppm) data. Nearly a decade ago, a new device called the Orbitrap mass analyzer [72] (Fig. 1.10 C) provided features of both high - resolution and high mass accuracy [73]. The latest generation of Orbitrap mass analyzer offers maximum resolution of up to 450,000 (FWHM at 200 m/z) and mass accuracy between 1 - 3 ppm (Orbitrap Fusion tribrid MS, Thermo Scientific).

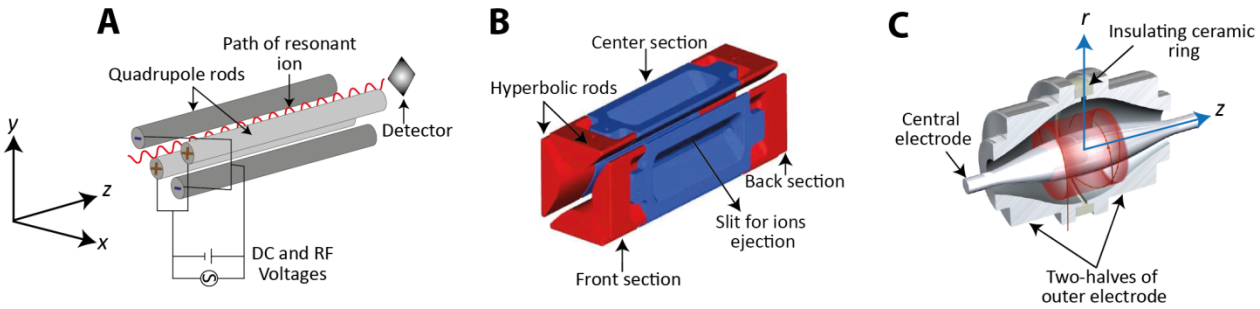


Figure 1.10: The working principles of the three commonly used MS in the field of proteomics. **(A) Quadrupole** mass analyzer consists of four precisely parallel metal rods. Opposite electrodes are connected and one pair receives a positive, the other pair a negative direct current (DC) potential that is superimposed by a time-dependent radio frequency (RF) potential. When ions are injected into the quadrupole in the direction of the rods, the oscillating electric field in the center of the quadrupole can be set to allow only a narrow mass-to-charge (m/z) range to pass on a stable trajectory. Stable oscillations are only achieved by ions of given m/z values for a given rod assembly, oscillation frequency, RF voltages, and DC voltage. The remaining ions will strike the rods. Thus, a quadrupole rather acts as a mass filter than as a conventional MS. Ramping the amplitude of the DC and RF potentials enables different narrow m/z ranges to pass the quadrupole and thereby generates a mass spectrum. **(B)** The design of the quadrupole **linear ion trap** (2D trap) by Jae Schwartz et al. [74] resembles a quadrupole that is split into three sections. The central section of the three parts is the largest and is intended to store the ions, whereas the front and back sections can be used for ion manipulation and for applying an axial trapping potential. Ions are trapped by DC potentials applied to ends of the four hyperbolic rods, X and Y (on all three segments of each rod) and RF applied to all rods. The behavior of the ions and their movement is explained by the *Mathieu's* differential equations and stability diagram (for a detailed explanation please refer to [75]). Ramping the amplitude of the main RF, ions leave the trap - low m/z to high m/z. An ejection slit in one of the central rods that allows ion ejection and detection by an electron multiplier. Besides storing ions, 2D traps can be combined with other mass analyzers in hybrid instruments and used to isolate ions of selected m/z ratios, to perform tandem mass spectrometry (MS/MS) experiments. **(C)** The **Orbitrap** mass analyzer is based on an earlier ion storage device, the Kingdon trap [76], which uses electrostatic fields to trap and analyze ions and consists of one central spindle and one outer barrel-like electrode that are connected by a ceramic ring. Ions injected into the Orbitrap are electrostatically trapped while rotating and oscillating along the central electrode. The axial oscillating frequency is dependent on the m/z of the ion. Oscillating ions induce an image current signal on the outer electrodes. Image current signals are converted into frequencies by Fourier transformation. The frequencies, which are characteristic of each ion m/z value, are finally converted into a mass spectrum [72].

To meet the requirements of the proteomics research area, which typically demands rapid and accurate analysis of biomolecules; hybrid MS were developed recently which combine two or more different types of mass analyzers. For example in the LTQ Orbitrap XL (Thermo Scientific), throughput and sensitivity are maximized by acquiring MS data at high resolution and accuracy with the Orbitrap, whereas MS/MS (see next chapter) data is recorded (in parallel or alternating fashion) at high speed with low resolution in a linear ion trap mass analyzer (Fig. 1.11). A brief

description of the three different hybrid mass spectrometers that were employed in this work is given below (Figs. 1.11 - 1.13).

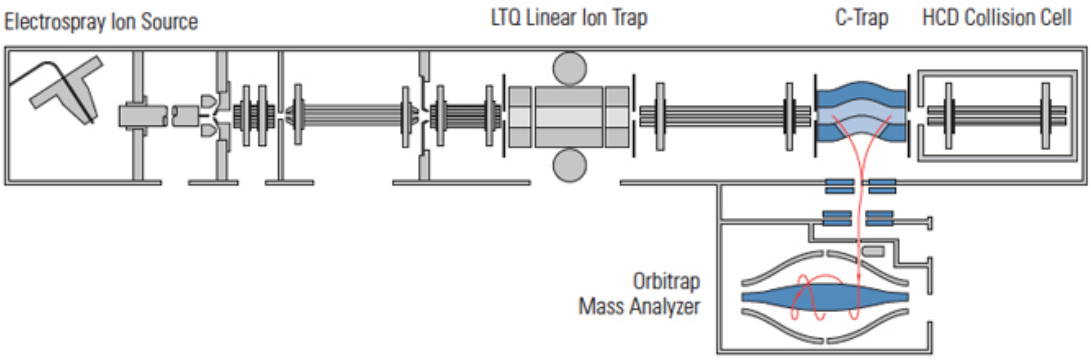


Figure 1.11: Schematic of a **LTQ Orbitrap XL** MS (Thermo Scientific, http://planetorbitrap.com/). The linear trap quadrupole (LTQ, known as linear ion trap) is used at the front end for ion trapping, ion selection, fragmentation reactions and low resolution ion detection, whereas the Orbitrap is used for high resolution (R > 100,000), high mass accuracy (< 3 ppm) ion detection [73].

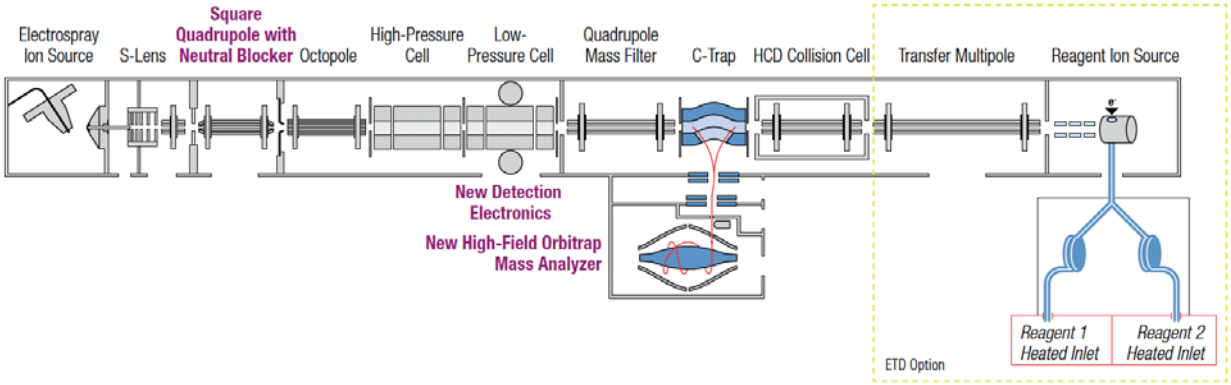


Figure 1.12: Schematic of an **Orbitrap Elite** MS (Thermo Scientific, http://planetorbitrap.com/). Key differences from LTQ Orbitrap XL are: the single linear ion trap was replaced by two identical linear ion traps that are operated at different gas pressure regimes. The first trap is held at higher pressure for improved trapping and fragmentation whereas the second trap is kept at lower pressure for enhanced scanning capabilities. Furthermore, the HCD collision cell was modified for faster extraction and more efficient transmission of all ions. Key improvements compared to the LTQ Orbitrap Velos (which has an identical instrument design) are: faster scan speed and better dynamic range of the linear ion trap and the novel "high-field" Orbitrap, which offers a resolution of 240,000 at m/z 400 at one scan/second (1 Hz) [77].

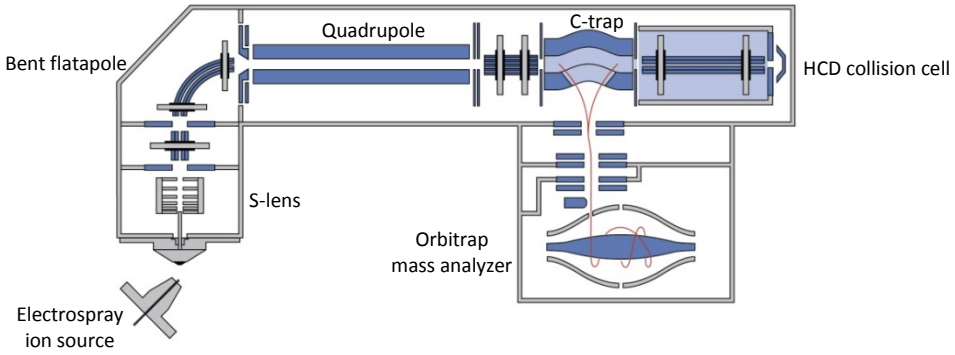


Figure 1.13: Schematic of a **Q Exactive** MS (Thermo Scientific, http://planetorbitrap.com/) with the Orbitrap for high resolution and high mass accuracy analysis of both precursor and product ions. Peptide ions are isolated for MS/MS by the quadrupole and are subjected to HCD fragmentation and stored in the C-trap, which allows for ion accumulation and Orbitrap mass analysis of ions from the preceding event in parallel, resulting in fast duty cycle. Furthermore, implementation of the enhanced Fourier Transform (eFT) algorithm to process the image current from the detector results in a twofold increase in resolution in comparision to the standalone Orbitrap analyzer called Exactive [78].

1.4.1.5 Tandem mass spectrometry

In bottom-up proteomics, tandem mass spectrometry (or MS/MS) analysis is performed to obtain amino acid sequence information of the peptide ions [79]. Peptide sequencing by MS/MS method involves at least two stages of mass analysis. In the first stage MS analysis, a peptide (or precursor) ions of a specific m/z window (e.g. 500.0 - 502.0 m/z) are selected and subsequently dissociated. Among different types of fragmentation procedures, collision induced dissociation (CID) is the most frequently employed peptide dissociation technique [80]. In CID, the selected precursor ion (m/z value) is activated by collisions with inert gas atoms (e.g. Argon, Helium or Nitrogen) [81]. With each collision, a part of the kinetic energy is converted into internal (vibrational/rotational) energy of the precursor ion. If the gained internal energy is high enough the precursor ion will dissociate resulting in the cleavage of the weakest bond first, which is usually the CO-NH amide bond in peptides thus generating predominantly the so-called b- and y-type fragment (product) ions [82] (Fig. 1.14). Fragmentation is a statistical process where certain amino acid (combinations) dissociate more easily than others, but owing to the huge numbers of peptide ions that are fragmented, many different types of b- and y-ions are generated that help to sequence a peptide. In the second stage of MS, the resulting m/z values of these fragment (or product) ions are analyzed.

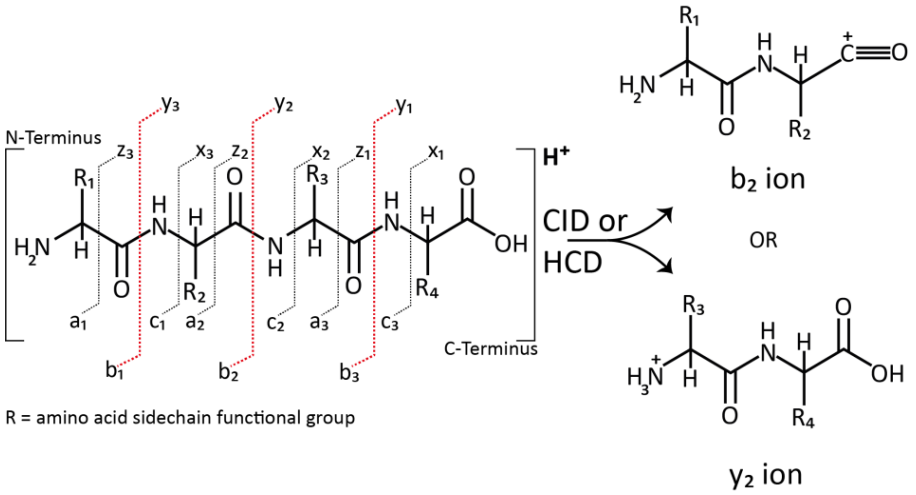


Figure 1.14: There are mainly three different types of bonds that can dissociate along the peptide backbone i.e. CH-CO, CO-NH and NH-CH bonds upon chemical reactions with inert gases. The nomenclature for the product ions as per the fragmentation rules is "a, b, c" ions containing the N-terminus and the "x, y, z" ions containing the C-terminus of the peptide, respectively [82-83]. Collisional-induced dissociation such as CID and HCD predominantly cleave the peptide bond (CO-NH) and generate b-type or y-type ions. In the figure, b_2 and y_2 ions are shown as representative structures of charged product ions. As many different types of b- and y-ions are generated during fragmentation, these ions can be used to sequence a peptide in an MS/MS scan.

The instruments that are capable of performing MS/MS experiments can be classified into two groups: tandem-in-space and tandem-in-time instruments [84]. Tandem-in-space instruments require separate mass analyzers to be utilized for each MS stage e.g. triple quadrupole mass spectrometer, wherein the first and last quadrupoles act as the actual mass filters with the middle quadrupole acting as a CID cell. The type of CID process performed by such instruments is referred to as beam-type CID. In contrast, tandem-in-time instruments separate the different MS stages by time with the various stages of MS/MS being performed in one mass analyzer e.g. linear ion trap. This type of CID process is known as resonant-excitation collision induced dissociation or ion trap CID. Here, fragmentation is achieved through hundreds of weak collisions, and ion trap CID has a certain limitation i.e. ions below 30% of the precursor m/z are lost during trapping making it unsuitable for the isobaric-tagging based quantitative techniques that rely on low m/z values. Nevertheless, linear ion trap - Orbitrap hybrid instruments e.g. Orbitrap XL, Velos or Elite, Q Exactive (Figs. 1.11 - 1.13) are equipped with a collision cell that can perform higher-energy collision induced dissociation (HCD), which is similar to the fragmentation achieved in linear quadrupole mass spectrometers i.e. beam-type CID [85]. In beam-type CID, fragmentation is achieved by single higher energy collisions and no low m/z value cutoff limitation is given, therefore it is well-suited for experiments e.g. isobaric tags for relative and absolute quantitation (iTRAQ) [86].

1.4.1.6 MS data acquisition strategies

Depending on the configuration and performance characteristics, the MS data can be acquired in several ways. There are two different approaches that are commonly used in the MS-based proteomics namely: (i) data dependent acquisition (DDA) and (ii) data independent acquisition (DIA), the latter will not be further explained since it was not applied in this work. In DDA mode, a full MS scan over a set m/z range (e.g. 300 - 1,500 m/z) is acquired at first. This represents the m/z of all ions (peptide) present at that time point. Next, typically the N most abundant precursor ions (Top N) present in that MS scan are selected, separately isolated and subjected to fragmentation and MS/MS analysis. After acquiring the N MS/MS spectra to complete the ''cycle'', the instrument repeats this procedure. For instance, in a Q Exactive MS, one cycle comprising a single full MS scan in the Orbitrap at a resolution of 70,000 (FWHM at 200 m/z) and ten (Top 10) DDA HCD MS/MS scans (Orbitrap at R = 17,500) takes around 1.1 s. To avoid redundant acquisition of MS/MS from the most abundant peptide ions in typical LC elution windows of ~5-20 s, the so-called dynamic exclusion is used, which usually puts the m/z value of a precursor ion into an exclusion list for a set duration of time after its MS/MS spectrum has been acquired once. Despite being a highly sensitive approach, it has been observed that the DDA is biased toward the most abundant peptides mainly due to (i) stochastic nature of the precursor ion selection and (ii) the extreme complexity of the sample (e.g. human cell lysate), which overwhelms the performance of LC separation and the scan speed of the instrument thus leading to undersampling of the low abundant species [87]. This consequently hampers the overall dynamic range of the analysis [88].

1.4.1.7 Database dependent protein identification

In this approach, peptide identification is performed by correlating experimental tandem MS spectra with theoretical spectra predicted for each peptide contained in a protein sequence database. The search database is simply a FASTA text file, where individual protein sequence entries are concatenated with the protein identifier as a delimiter, with alternative protein isoforms being handled as an individual protein entry. Many different algorithms have been developed for identifying tandem MS data using database search engines, which operate in a

similar manner and follow the same general workflow. The acquired MS/MS spectra are compared and correlated against theoretical spectra constructed for each database entry that satisfies a certain set of database search parameters i.e. mass tolerance (which depends on the MS system that was used), enzyme constraint, and types of PTMs specified by the user. A scoring scheme is then used to measure the degree of similarity between the experimental MS/MS spectra and theoretical fragmentation patterns. Next, the candidate peptides are ranked according to the computed score and the highest scoring peptide sequence (peptide to spectrum match or PSM) is selected for further analysis (Fig. 1.15). The identified peptide sequences are then mapped to their corresponding proteins followed by statistical evaluation and validation [61]. The main difference between different search algorithms is the scoring function used to quantify the degree of similarity between the acquired tandem mass spectrum and the candidate peptides retrieved from the database. For instance Sequest [89] scores peptide sequences by the cross-correlation between the intensities of peaks on the observed and the theoretical spectra whereas, Mascot [90] scoring is based on the absolute probability that the observed match is a random event. An ion score is reported as $-10*\log_{10}(p)$, where p is the absolute probability. A higher score indicates a more confident match. Furthermore, to assess the reliability and validate the reported PSMs, a

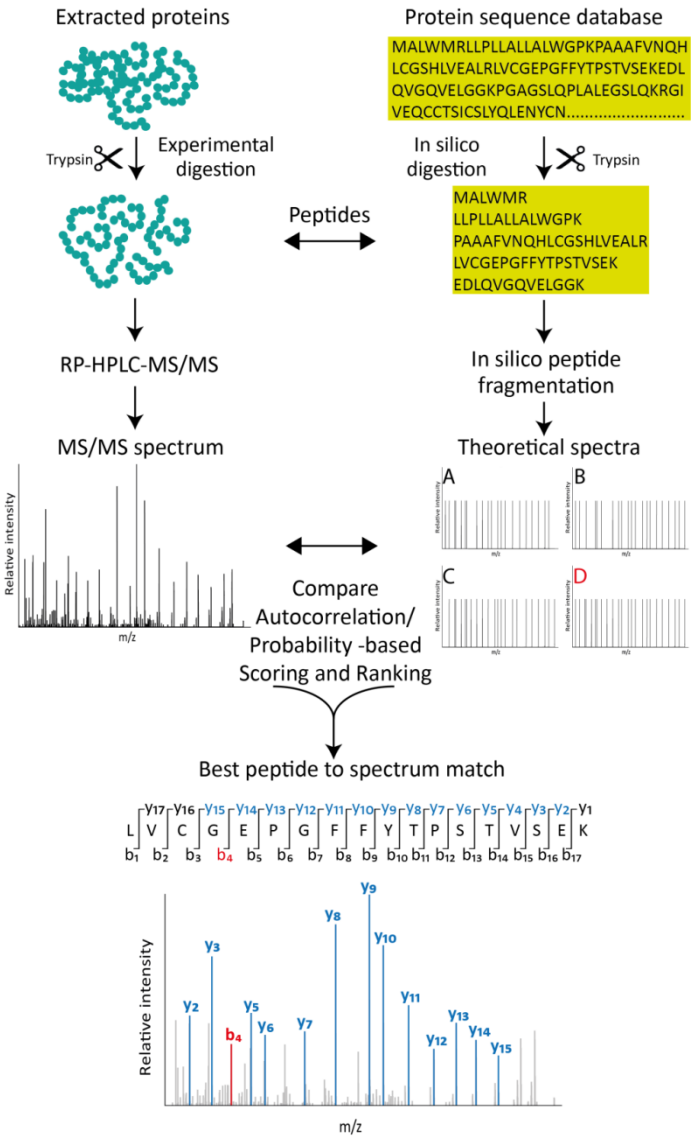


Figure 1.15: Peptide identification by correlating acquired MS/MS spectrum with theoretical spectra predicted for each peptide contained in a protein database. Most search algorithms mainly use m/z information (e.g. all the y ions are given equal relative abundance in the theoretical spectra) and not intensity information. In the figure, the best PSM is D (in red color).

target/decoy search strategy was introduced [91]. In this approach, the MS/MS spectra are searched by algorithms not only against the standard sequence database (target), but also against a database containing usually reversed protein sequences (decoy). The idea is that PSMs obtained from the decoy database are random hits and therefore can be used to estimate the number of incorrect (random) target PSMs for any given criteria such as score thresholds or heuristic methods. This enables the calculation of the false discovery rate (FDR) by simply counting the number of decoy and target PSMs that meet the chosen acceptance criteria [92]. Usually, a 1% FDR is used on PSM, peptide or protein level as a threshold for high confident protein identifications.

1.4.2 Quantitative proteomics to study human diseases

Quantitative proteomics is a prerequisite to investigate the highly dynamic proteome of living organisms as it describes the differences in expression of proteins among different biological states (e.g. healthy vs. disease, wild type vs. specific mutation) [93]. However, MS is not *per se* "inherently quantitative" [57a] since peptides generated from intact proteins, each of which has unique physical (e.g. MW, amino acid composition) and chemical (e.g. hydrophobicity, pI) properties will influence the overall LC-MS performance characteristics that might complicate peptide/protein quantification [57a, 94]. For this reason, stable (i.e. non-radioactive) isotope based strategies have been introduced that take the advantage of a MS ability to quantitatively distinguish the relative abundances of these heavy isotopes in otherwise identical chemical species, regardless of competing ion concentrations [95]. The basic assumption of stable isotope labeling for peptide and protein quantification is that the physicochemical properties of the differentially labeled peptides are nearly identical [96]. This includes sample preparation procedures, LC-separation performance, ionization efficiency and MS/MS fragmentation behavior [97]. The most commonly used stable isotopes are: ^{13}C, ^{15}N, ^{2}H, ^{18}O and they can be introduced by (i) chemical reactions of a labeling reagent with a distinct functional group (e.g. iTRAQ) [98], (ii) metabolic processes (e.g. stable isotope labeling by amino acids in cell culture, SILAC) [57b] or (iii) enzyme (e.g. trypsin)-facilitated ^{18}O labeling (e.g. $H_2{}^{18}O$) [99]. Depending on the type of label, differentially labeled samples can be combined at different steps of the sample preparation workflow assuming an overall near-complete labeling efficiency. Peptide

quantification is performed either on MS or MS/MS level. The basis of quantification is the shift in absolute mass corresponding to the stable isotope incorporated into the peptide or protein. Thus in case two samples were labeled either "light" or "heavy", pooled and analyzed together, typically each peptide will appear in both, the light and the heavy form in the final LC-MS analysis. For each of these two forms of the peptide the area under the curve of the isotopic envelope is integrated over the LC elution time. Finally, as the units obtained in MS are rather arbitrary and do not directly allow deducing peptide concentrations/amounts - a relative quantification, i.e. determining the relative change across samples, is performed by calculating the ratio of the peak areas of the differentially labeled peptides. In so-called label-free precursor ion intensity-based quantification works with the same principle however, peptide abundances are retrieved from consecutive LC-MS analyses since owing to the lack of stable heavy isotopes, each sample has to be analyzed individually. Moreover, targeted MS methods such as selected reaction monitoring (SRM) and parallel reaction monitoring (PRM), which offer the specific analysis of dedicated proteins - comparable to immunoblot assays (e.g. WB) may be used for protein quantitation [100]. In the present work, chemical labeling (by iTRAQ reagents), label-free approaches and PRM were employed for relative protein quantification and their principles and applications are given below.

1.4.2.1 Protein quantification with labeling reagents

The use of isobaric mass tags to monitor relative changes in protein abundances across altered biological systems have been part of LC-MS-based proteomics for more than a decade. iTRAQ [98] is one of the most popular technologies and it enables relative protein quantification of up to 8 distinct biological samples in a single LC-MS/MS analysis. An iTRAQ label consists of a primary amine reactive group, a balancer group, and a reporter group. The amine-reactive, (N-hydroxysuccinimide ester) group reacts mainly with the unblocked (e.g. acetylated) primary amines present at the N-termini of proteins/peptides and the epsilon side chain of Lys residues to attach the tags to the respective biomolecules. The mass normalization group (carbonyl moiety) compensates for the mass difference among the different reporter ion (N-methylpiperazine) groups such that different isotopic variants of the tag are isobaric. The overall mass of reporter and balance groups of the reagent are kept constant using differential isotopic

enrichment with ^{13}C, ^{15}N, and ^{18}O atoms. In an iTRAQ 4-plex set, the reporter group ranges in nominal mass from m/z 114 - 117, whereas the balance group ranges in mass from 28 to 31 Da, so that the combined mass remains constant (145 Da) for each of the four reagents (Fig. 1.16 A).

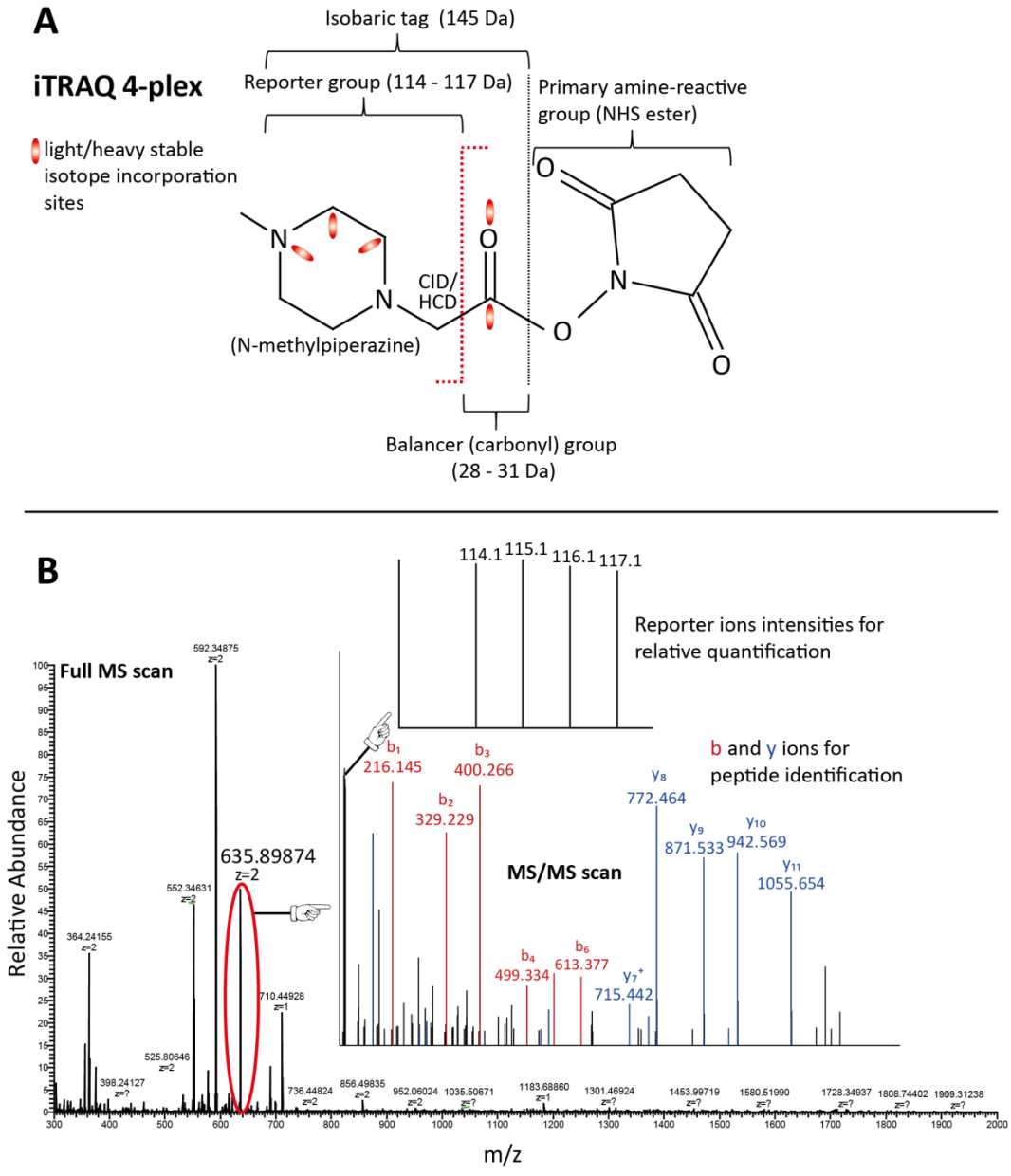


Figure 1.16: (**A**) Chemical structure of an iTRAQ 4-plex reagent. (**B**) Representative MS and MS/MS spectra of an iTRAQ labeled peptide (4-plex) which was subjected to HCD fragmentation. The full MS scan shows a single precursor ion peak isolated for HCD. The MS/MS spectrum depicts a series of product ions (b- and y- type) which are used for peptide identification, whereas the reporter ion signals in the low m/z region are used for relative quantification.

To increase the multiplexing capacity, iTRAQ 8-plex version was introduced with reporter ion masses at m/z 113 - 118, 119 and 121 (and balance groups ranging from 24 to 31 Da). Typically, iTRAQ labeling is performed on the peptide level after proteolytic digestion. Each peptide mixture is labeled with a different channel of iTRAQ reagents. The differentially labeled samples are then combined and subjected to LC-MS/MS analysis. Due to the isobaric nature of the tags, the differentially labeled peptides are indistinguishable during the full MS scan and are consequently jointly selected for MS/MS. Upon fragmentation, each tag releases its individual reporter ion and the signal intensities of these ions then reflect the peptide abundances between differently labeled samples. In parallel, the sequence of the peptide is determined from the product ions (b- and y-type ions) that are generated by the cleavage of peptide bond in the same MS/MS spectrum (Fig. 1.16 B).

There are several advantages of iTRAQ compared to other strategies. Firstly, multiplexing, which is the ability to combine and analyze several samples (up to 8) within one experiment eliminates the need to compare multiple LC-MS/MS datasets and can reduce run-to-run variation. Secondly, as the chemical derivatization process is performed on the peptide level, iTRAQ allows labeling of virtually any samples, including mammalian tissues and body fluids such as blood serum/plasma samples, which is not feasible with metabolic labeling (e.g. SILAC). Thirdly, due to the isobaric nature of the tags, multiplexing does not increase sample complexity which is the case for SILAC and the signal intensity of the same peptide from all samples is summed up during the full MS scan. This increases the sensitivity of the analysis and allows identification and quantification of low-abundant proteins in complex samples [101]. However, similar to other reporter ion based quantification strategies, iTRAQ is prone to the so-called "ratio compression" issue, which occurs due to the interference from interfering peptide ions that are within the isolation window (typically 2 - 3 m/z) [102]. As all peptides release the same reporter ions such co-isolation interfering peptides results in distortion of the reporter ion intensities of the peptide of interest. Furthermore, reporter ion interference by co-isolated contaminating peptides can incline the observed ratios towards unity as the expression of majority of the proteins remains unchanged in most experiments. This leads to an underestimation or compression of actual protein abundance differences in the analyzed samples and consequently affects overall quantification accuracy [103]. To address this interference issue, several solutions have been

proposed. By employing a robust fractionation step prior to LC-MS analysis could solve the co-eluting peptides problem by reducing the sample complexity [104] and narrowing the precursor ion isolation window during MS analysis might reduce the amount of undesired interference [105]. Recently, Ting et al. proposed triple-stage mass spectrometry (MS3) approach to overcome the reporter ion interference problem more efficiently. In this method isolated peptide ion is initially subjected to ion trap-CID fragmentation. Subsequently, one of the most intense fragment ions is isolated for HCD fragmentation to generate the reporter ion intensities [106].

1.4.2.2 Label-free protein quantification

Label-free MS methods for relative protein quantification of different biological states or to estimate the absolute protein abundances in a given cell type are a promising alternative over labeling based strategies [107]. This is mainly because (i) it requires less sample preparation steps, (ii) they can be applied to any biological material, (iii) the complexity of the sample is not increased since mixing of different proteomes is not required, and therefore (iv) no ratio compression can occur. MS-based label-free protein quantification can be performed in two ways i.e. using precursor ion areas/intensities or by spectral counting [108] (Fig. 1.17).

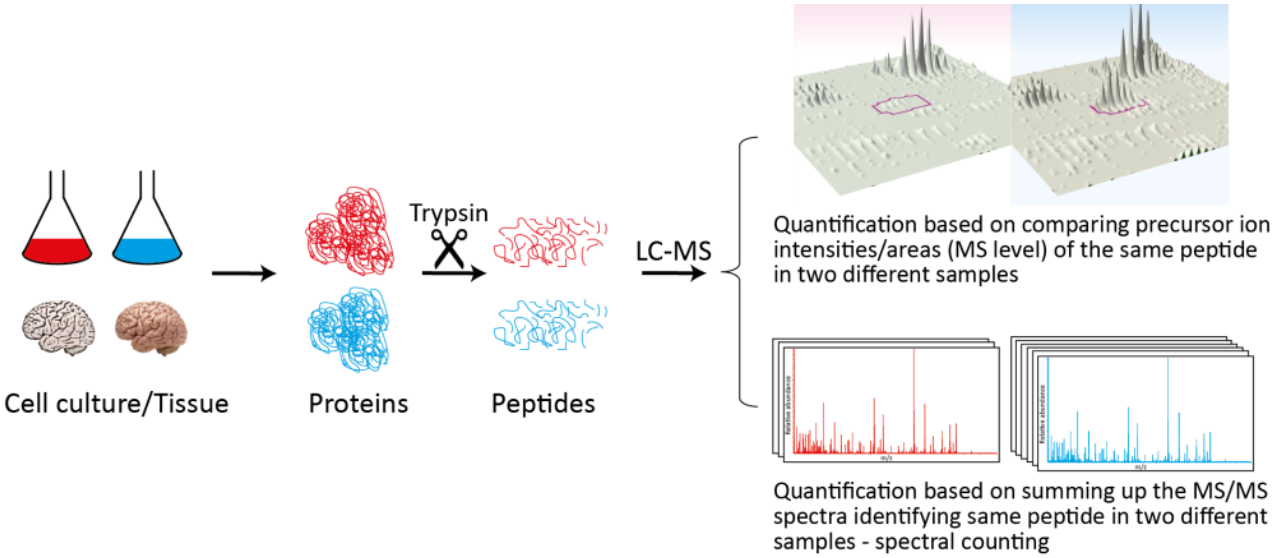


Figure 1.17: Proteins extracted from either cell culture/tissues that belong to two different conditions (in the figure red and blue colors represent disease and healthy states, respectively) are enzymatically digested and analyzed directly by LC-MS. Label-free relative protein quantification can be done by using precursor ion intensities (top) or spectral counting (bottom). In the above hypothetical example, the same peptide in blue (or healthy) sample is more abundant relative to the red (or disease) sample.

Precursor ion intensity-based quantification relies on the principle that the area of a peak in the full MS scan is a measure of the abundance of the corresponding peptide in the sample i.e. a twofold increase in this signal should reflect a twofold increase of this peptide. Peptides are identified based on their MS/MS spectra and then the corresponding precursor peaks are identified in each LC-MS run. Depending on the used software, the areas under these peaks are calculated and the areas of different peptides that belong to same protein are summed. By comparing these summed peak areas across different samples, relative protein quantification is achieved. For instance, all the steps that are involved in this type of label-free quantification can be performed using dedicated software e.g. Progenesis (see Section 3.2.14.2.1). In contrast, the spectral counting approach relies on the MS/MS data acquired from the DDA experiments. Relative quantification is done by comparing the sum of PSM for a given protein across multiple samples [107]. The rationale is that increasing amounts of protein/peptide will lead to higher numbers of PSMs. Furthermore, the spectral counting approach can also be extended for the rough estimation of protein abundances in a given sample. The frequently used methods for this purpose in bottom-up proteomics experiments are: exponentially modified protein abundance index (emPAI) [109], absolute protein expression index (APEX) [110] and normalized spectral abundance factor (NSAF) [111]. Recently, McIlwain and co-workers performed a comparative analysis between different spectral counting methods which are currently used for estimating relative protein abundances in complex proteomes. Their study concluded that NSAF performs better in terms of linear response to protein abundance and gives more reproducible results when compared to other methods e.g. emPAI [112]. In this work, the NSAF approach was used to estimate the protein abundances in human RCMH cell line and to perform relative protein quantification of human skeletal muscles (see Section 3.2.14.2.2). Despite being a straightforward approach, the major drawback of all label-free quantification methods is the massive increase in total analysis time as each sample is measured individually. Consequently, the probability of variation between the LC-MS runs of different samples might increase leading to decrease in precision and accuracy. Thus, label-free quantification approaches require extremely robust instrumentation that is capable of providing reproducible LC separation and subsequent MS detection [57a]. However, in DDA-based label-free experiments, the overall

protein quantification is hampered due to its inherent drawbacks i.e. stochastic nature of precursor ion selection and bias towards high abundant peptides.

1.4.2.3 Protein quantification with targeted MS

Selection reaction monitoring (SRM) and the recently introduced parallel reaction monitoring (PRM) [113] are the commonly used MS-based technologies in the field of targeted quantitative proteomics [114]. Owing to their non-scanning nature (i.e. no full MS is recorded), both sensitivity and linear response over a wide dynamic range are increased when compared with other MS data acquisition techniques such as DDA and DIA. This enables the detection of low-abundance proteins in highly complex mixtures, which is crucial for systematic quantitative studies [115].

SRM is usually performed on a triple quadrupole instrument wherein the first and the third quadrupoles act as filters to specifically select predefined m/z values corresponding to the precursor ion (Q1) and a specific product ion of the peptide (Q3), respectively. The second quadrupole (Q3) serves as a collision cell for fragmentation (beam type CID). Only if the correct pair of precursor and product ion m/z values, referred to as a "transition" is present, a signal is detected. In SRM, several such transitions per peptide are monitored and their peak areas are integrated, similar to label-free quantification described above. The difference is, that here not the entire sample is quantified, but pre-defined, selected peptides, thus allowing a highly sensitive, specific and fast analysis with highest precision [100] (Fig. 1.18 A). Whereas, in the PRM approach, the last quadrupole (Q3) is replaced with a high resolution mass analyzer (e.g. the Orbitrap) to acquire a full MS/MS. Quantification in PRM is carried out by extracting several fragment ions of the targeted peptide post-acquisition on the software level as "pseudo-transitions", which are integrated to generate corresponding peak areas [100, 113] (Fig. 1.18 B). However, PRM-based targeted analysis is emerging as a highly selective and sensitive approach for protein quantification since the full MS/MS spectra of the targeted peptide ions are acquired with high resolution e.g. 70,000 (FWHM at m/z 200) and high mass accuracy (< 1 ppm) when compared to SRM that offers both low resolution and mass accuracy data generated by the quadrupole (R = 2,000 and ~100 ppm, respectively). Additionally, as all the product ions of a targeted peptide are detected during PRM analysis, there is no essential requirement for the

product ion m/z values information (which is a prerequisite for SRM), which saves time in the assay method development [113, 116].

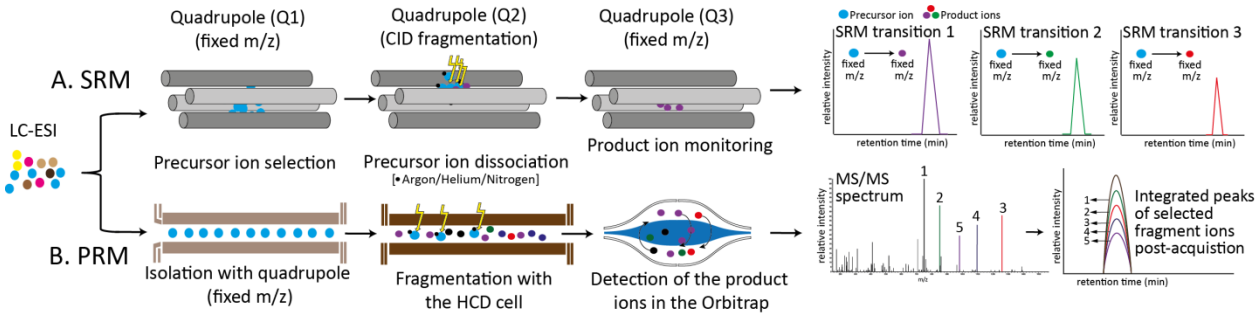


Figure 1.18: (A) Schematic of SRM performed on a triple quadrupole mass spectrometer. In SRM, the first quadrupole (Q1) is used for isolation and the second (Q2) as a collision chamber for fragmentation and the third quadrupole for mass selective transfer of fragment ion peaks, to generate so-called precursor-fragment transitions. **(B)** Schematic of PRM performed on a Q Exactive mass spectrometer (figure adapted from Thermo Scientific). In PRM, a target precursor ion is isolated by the quadrupole analyzer with a narrow m/z window (up to ±0.2 m/z) and fragmented in the HCD cell. The resulting fragment ions are then co-detected in the Orbitrap mass analyzer to generate high resolution and high mass accuracy MS/MS data. Peak areas of fragment ions are extracted using narrow mass (≤ 10 ppm) windows and integrated across the chromatographic elution profile. During the data processing, subsets of fragment ions with highest intensities in the MS/MS spectrum are used for both peptide identification and quantification using dedicated software e.g. Skyline [117].

In this work, PRM was used to establish an assay for quantifying UPR-related proteins (see Section 1.2). As some of the UPR-related proteins/factors are usually present in low abundance in a complex cell lysate, a PRM-based targeted assay is an ideal method for the precise identification and quantification of these proteins. The details of the assay method development are given in the sections 3.2.13 and 3.2.15.3.

2 Aim

The goal of this work was to study pathological implications caused due to altered protein folding mechanisms in the endoplasmic reticulum (ER). For this, quantitative proteomics was used to study Marinesco-Sjögren Syndrome (MSS; OMIM 248800) as a model. MSS is a rare, genetically inherited multi-systemic disorder and the patients mainly suffer from the eyes, the brain and skeletal muscle abnormalities. Genetic studies have revealed mutations in *SIL1*/*Sil1* as the main cause of MSS in man and mouse, respectively. SIL1 is a nucleotide exchange factor (NEF) for the ER-resident chaperone BiP, which controls a plethora of essential processes and most importantly assists in (un)folding of proteins in the ER. Loss of functional SIL1 compromises the folding capacity of BiP leading to the aggregation of misfolded proteins and eventually failure of cellular functions. Despite being ubiquitously expressed, loss of SIL1 affects only certain tissues/organs and several questions remained unsolved pertaining to the selective vulnerability of SIL1 loss. The existence of other rescue mechanisms besides the presumed overexpression of GRP170 - an alternate NEF for BiP, is speculated in the SIL1 non-vulnerable cells/tissues, albeit these processes have not been identified so far. The function of SIL1 as a key player in the protein folding process in the ER and consequently its impact on cellular homeostasis renders quantitative proteomics as promising tool to understand dysfunctional mechanisms and compensatory mechanisms associated with SIL1.

Therefore, the main aims of this dissertation were (i) to identify possible alternate mechanisms that compensate for the loss of functional SIL1 and thus maintain homeostasis in SIL1-deficient unaffected tissues and (ii) to see the global impact of SIL1 loss on the proteome level in SIL1-deficient affected tissues using MS-based proteomics. Moreover, to establish quantitative workflows to study genetic defects e.g. MSS on the level of primary cell cultures, tissue biopsy and *in vitro* models (e.g. HEK293). Additionally, the role of the unfolded protein response (UPR) pathway in maintaining cellular homeostasis due to the impaired SIL1-BiP protein folding machinery was briefly addressed.

3 Materials and methods

3.1 Materials

3.1.1 Chemicals

Table 3.1: Used chemicals and reagents.

Name	Abbreviation	Formula	Supplier	Place, Country
2-Mercaptoethanol		C_2H_6OS	Sigma Aldrich	Steinheim, Germany
β-casein, β-lactoglobulin, haemoglobin, serum albumin (all Bovine) and horse myoglobin			Sigma Aldrich	Steinheim, Germany
Acetone		C_3H_6O	Merck KgaA	Darmstadt, Germany
Acetonitrile	ACN	C_2H_3N	Biosolve BV	Valkenswaard, the Netherlands
Ammoniumacetate		$C_2H_3O_2NH_4$	Fluka, Sigma Aldrich	Steinheim, Germany
Ammoniumbicarbonate		NH_4HCO_3	Fluka, Sigma Aldrich	Steinheim, Germany
Ammoniumhydroxide solution, 25% (v/v)		NH_4OH	Fluka, Sigma Aldrich	Steinheim, Germany
Benzonase Nuclease, Purity > 99%			Merck Chemicals GmbH	Darmstadt, Germany
Bicinchoninic acid assay kit	BCA		Pierce, Themo Scientific	Rockford, USA
Calcium chloride		$CaCl_2$	Merck KgaA,	Darmstadt, Germany
complete Mini EDTA-free			Roche Applied Science	Penzberg, Germany
Dithiothreitol	DTT	$C_4H_{10}O_2S_2$	Roche Diagnostics GmbH	Mannheim, Germany
Ethanol	EtOH	C_2H_5OH	Merck KgaA	Darmstadt, Germany
Formic acid	FA	CH_2O_2	Biosolve BV	Valkenswaard, the Netherlands
Guanidine hydrochloride	GuHCl	CH_6ClN_3	Sigma Aldrich	Steinheim, Germany
Glycerol			Merck	Hohenbrunn, Germany
Iodoacetamide	IAA	C_2H_4INO	Sigma Aldrich	Steinheim, Germany
iTRAQ Reagent-4/8Plex Multiplex kits			AB SCIEX	Framingham, USA
Isopropanol		C_3H_8O	Biosolve BV	Valkenswaard, the Netherlands
Kasil			PQ Silicas BV	Eijsden, The Netherlands
Lys-C (Sequencing Grade)			Roche Diagnostics GmbH	Mannheim, Germany
Magnesium chloride (anhydrous)		$MgCl_2$	Sigma Aldrich	Steinheim, Germany
RapiGest SF			Waters Corporation	Milford, MA, USA
Sodium chloride		NaCl	Merck KgaA	Darmstadt, Germany
Sodium dodecylsulfate	SDS	$NaC_{12}H_{25}SO_4$	Carl Roth GmbH & Co.	Karlsruhe, Germany
Sodium deoxycholate		$C_{24}H_{39}NaO_4$	Sigma Aldrich	Steinheim, Germany
Sodium hydroxide		NaOH	Merck KgaA	Darmstadt, Germany
Thiourea		CH_4N_2S	Sigma Aldrich	Steinheim, Germany
Triethylammonium bicarbonate	TEAB	$C_7H_{17}NO_3$	Sigma Aldrich	Steinheim, Germany
Tris(hydroxymethyl)aminomethane base	Tris	$C_4H_{11}NO_3$	Applichem GmbH	Darmstadt, Germany
Triton X-100			Roche Diagnostics GmbH	Mannheim, Germany
Trifluoroacetic acid	TFA	CF_3CO_2H	Biosolve BV	Valkenswaard, the Netherlands
Trypsin (T-1426)			Sigma Aldrich	Steinheim, Germany
Trypsin (Sequencing Grade Modified) and Trypsin Gold			Promega	Madison, WI 53711, USA
Urea		CH_4N_2O	Sigma Aldrich	Steinheim, Germany

3.1.2 Instruments and disposable consumables

Table 3.2: Used laboratory equipment and instruments.

Name	Supplier	Place, Country
3100 OFFGEL Fractionator	Agilent	Waldbronn, Germany
96-well plates (0.3/0.5/1.0/2.0 mL)	Eppendorf	Hamburg, Germany
96-well plate covering foils (RNase/DNase-free, not sterile)	Capitol Science	Austin, USA
Analytical balance (Cubis)	Sartorius Lab Instruments GmbH & Co.KG	Goettingen, Germany
Biological safety cabinet (HERASAFE KSP18), Built in UV-Lamps	Thermo Electron LED GmbH	Langenselbold, Germany
Cellstar tubes (15/50 mL)	Greiner Bio One	Frickenhausen, Germany
Centrifuge (5417 R, 5424, 5810 R, MiniSpin plus)	Eppendorf	Hamburg, Germany
Heating compartments	Memmert	Schwabach, Germany
Immobiline DryStrips, pH 3-10	GE Healthcare	USA
Microcentrifuge tubes (Protein LoBind: 0.5 mL, 1.5 mL, 2 mL, 5 mL)	Eppendorf	Hamburg, Germany
Microcentrifuge tubes (Safe-Lock: 0.5 mL, 1.5 mL, 2 mL)	Eppendorf	Hamburg, Germany
Microcon-30kDa Centrifugal Filter Unit with Ultracel-30 membrane	Merck Chemicals GmbH	Darmstadt, Germany
Microtiterplate reader (Multiskan FC)	Thermo Scientific	Bellefonte, USA
Mortar and Pestle (Agate/Achat: 10 mL)	VWR International GmbH	Darmstadt, Germany
NanoDrop 2000 UV-Vis Spectrophotometer	Thermo Scientific	Wilmington, USA
Nanosep Centrifugal Devices with Omega Membrane - 30kDa	Pall GmbH	Dreieich, Germany
OMIX C18 pipette tips (10 µL and 100 µL)	Agilent	Waldbronn, Germany
pH-electrode (Blueline14)	Schott Instruments GmbH,	Mainz, Germany
pH-test strips (2-3/6.0-7.7/6.0-10/7-14)	Macharey Nagel GmbH & Co.	Düren, Germany
Pipettes (0.5-2.5; 0.5-10; 2-20; 10-100; 20-200; 100-1000;1000-5000 µL)	Eppendorf	Hamburg, Germany
Pipette tips (0.5-2.5; 0.5-10; 2-20; 10-100; 20-200; 100-1000;1000-5000 µL)	Eppendorf	Hamburg, Germany
SPEC C18 AR (4 mg and 15 mg bed columns)	Agilent	Waldbronn, Germany
Thermomixer (0.5 mL, 1.5 mL, 2 mL)	Eppendorf	Hamburg, Germany
Ultrasonic bath (Sonorex)	Banderlin Electronic	Berlin, Germany
Ultrasonic processor (Vibracell 75022)	ACIL Sarl	Chatou, France
Vacuum centrifuge (Savant SPD 121P), RV6 vacuum pump, RVT 4104 cooling trap	Eppendorf	Hamburg, Germany
Vacuum centrifuge (RVC-2-18), RV6 vacuum pump, Beta 2-4 LP plus LT cooling trap	Martin Christ Gefriertrocknungsanlagen GmbH	Osterode am Harz, Germany
Vortex mixer (Genie-2)	Scientific Industries	New York, USA
Water purification system	ELGA LabWater	Celle, Germany
Vacuum manifold system	Agilent	Waldbronn, Germany

3.1.3 LC-columns, HPLCs and mass spectrometers

Table 3.3: Used columns, HPLC systems and mass spectrometers.

Name	Supplier	Place, Country
Acclaim PepMap C18 75 µm ID, 15/25/50 cm length, 3 µm particle size, 100 Å pore size	Thermo Scientific	Bellefonte, USA
Acclaim PepMap C18 100 µm ID x 2 cm length, 5 µm particle size, 100 Å pore size	Thermo Scientific	Bellefonte, USA
BioBasic C18 0.5 mm ID, 15 cm length, 5 µm particle size, 300 Å pore size	Thermo Scientific	Bellefonte, USA
Kinetex C18 2.6 µm particle size, 100 Å pore size (bulk material)	Phenomenex	Aschaffenburg, Germany
PepSwift Monolithic trap column, 200 µm ID, 5 mm length	Thermo Scientific	Dreieich, Germany
PepSwift Monolithic capillary column, 200 µm ID, 5 cm length, PS-DVB	Thermo Scientific	Dreieich, Germany
ZORBAX 300SB C18 0.5 mm ID, 15 cm length, 5 µm particle size, 300 Å pore size	Agilent	Waldbronn, Germany
UltiMate U3000 HPLC	Thermo Scientific	Germering, Germany
UltiMate U3000 nano Rapid Separation Liquid Chromatography (RSLC) HPLC	Thermo Scientific	Germering, Germany
LTQ Orbitrap Elite	Thermo Scientific	Bremen, Germany
LTQ Orbitrap Velos Pro	Thermo Scientific	Bremen, Germany
LTQ Orbitrap XL	Thermo Scientific	Bremen, Germany
Q Exactive and Q Exactive Plus	Thermo Scientific	Bremen, Germany

3.1.4 Data analysis software

Table 3.5: Used software for LC-MS data acquisition, control and analysis.

Software	Version(s)	Manufacturer/ Source
Chromeleon	6.8 (SR 8-11)	Thermo Fisher Scientific,Germering, Germany
Mascot	2.4	http://www.matrixscience.com
MS GF+		http://proteomics.ucsd.edu/software-tools/ms-gf/
Ontologizer	2.1	http://compbio.charite.de/contao/index.php/ontologizer2.html
OMSSA	2.1.9	http://pubs.acs.org/doi/abs/10.1021/pr0499491
PeptideShaker	0.28.0; 0.29.1; 1.0	http://compomics.github.io/projects/peptide-shaker.html
Progenesis	4.1	http://www.nonlinear.com
Proteome Discoverer	1.3; 1.4	Thermo Scientific, Bremen, Germany
ProteoWizard	2.2.2954	http://proteowizard.sourceforge.net
SearchGUI	1.14.4; 1.18.4; 2.0	http://compomics.github.io/projects/searchgui.html
Sequest		http://fields.scripps.edu/sequest
Skyline	2.5.0.6157	https://skyline.gs.washington.edu
STRING database	10	http://string-db.org
Xcalibur	2.2	Thermo Scientific, Bremen, Germany
X! Tandem	Jackhammer (2013.06.15)	http://www.thegpm.org/tandem

3.2 Methods

3.2.1 Cell and tissue samples for proteomics analyses

All human cell culture i.e. primary dermal fibroblasts, Epstein Barr virus (EBV) - transformed lymphoblastoid cells (LCs), SIL1-depleted HEK293, RCMH and tissue i.e. human (index patient), mouse (cerebella and skeletal muscles) samples for various proteomics analyses were prepared and provided (as denatured cell lysates, dissected and frozen tissues) by Dr. Andreas Roos and his group members at the Institute of Neuropathology, RWTH Aachen University Hospital, Aachen, Germany (Dr. Roos present address: Institute of Genetic Medicine, Newcastle University, United Kingdom). Furthermore, all biochemical experiments including cell viability assays, immunohistochemistry (IHC), plasmid transfections, tandem affinity purification (TAP) assay, Western Blotting (WB) and transmission electron microscopy (TEM) were performed under the supervision of Dr. Roos in the Aachen University Hospital. All research activities were approved by the Ethical Committee of the University Hospital RWTH (ethics approval number: EK104/10), Aachen and they were conducted in accordance with the 1989 declaration of Helsinki. Moreover, all procedures involving mice were approved by the University Hospital Aachen Institutional Animal Care and Use Committee and conducted in compliance with the legal standards for the care and use of laboratory animals. Because the genetic diagnosis was

performed before the patients underwent a biopsy, skeletal muscle specimens (clinically affected due to *SIL1* mutations) were not available for proteomics studies due to ethical reasons. Nevertheless, the MSS patients agreed to provide skin biopsies or peripheral blood for scientific purposes. Therefore, by making use of the skin biopsies and the blood withdrawals, fibroblast and immortalized lymphoblastoid cell lines (see below) were generated, respectively.

3.2.1.1 Clinically unaffected cell types

Five male MSS patients with genetically and biochemically proven *SIL1* mutations were included in this work (i.e. MSS2, MSS24, MSS32, MSS33, MSS94) [41a] along with respective healthy donors; matched for both age and sex. All patients presented with the major clinical hallmarks of SIL1-related MSS. Skin biopsies of MSS2 patient and control were dissected under sterile conditions and isolated primary human dermal fibroblast cells were cultured in FibroGRO™ (Millipore) culture medium supplemented with 10% Penicillin-Streptomycin. Lymphoblastoid cells (LCs) of the remaining four MSS patients and the respective healthy controls were obtained by peripheral blood withdrawal and were subsequently immortalized using the EBV-producing marmoset B-cell line (B95-8) as a source of EBV for stimulation and transformation. Culturing of the immortalized LCs was maintained at 37°C in a 5% CO_2 atmosphere.

3.2.1.2 SIL1-depleted human embryonic kidney 293 cells

Human embryonic kidney 293 (HEK293) cells were transfected with 20 µg of a mixture of human 29mer SIL1 HuSH short hairpin -RNA-plasmids (OriGene) using Lipofectamine 2000 (Invitrogen) according to the manufacturer´s protocol. 24 hours post-transfection, cells were split on three 10 cm plates and selection for stable transfected cells was carried out using 7.5 µg/mL puromycin (Millipore) as a selection antibiotic. Stably transfected clones (cell colonies) were harvested, cultured and tested for remaining SIL1 levels using immunoblot technique. Two SIL1-depleted cell clones (ΔSIL1_1 and ΔSIL1_2) as well as one scrambled (Scr) transfected cell clone was selected for subsequent proteomics, biochemical and morphological studies. Culturing of the HEK293 cells was maintained at 37°C in a 5% CO_2 atmosphere.

3.2.1.3 Human myoblastic RCMH cells

After defrosting, RCMH cells [118] were cultured in Dulbecco's modified Eagle's medium F-12-Ham containing 0.1% sodium bicarbonate (Sigma-Aldrich, Taufkirchen, Germany) and 12.5% fetal calf serum (Biowest) at 37°C in a 5% CO_2 atmosphere to a confluence of approximately 70% before subjecting to proteomics and morphological studies.

3.2.1.4 Clinically affected tissues

3.2.1.4.1 Mouse

Sil1 (located on murine chromosome 18) mutants were obtained from the Jackson Laboratories (strain name: CXB5/By-Sil1wz/J; stock number 003777). Homozygous affected animals (*woozy*) as well as wildtype littermates were obtained by mating heterozygous males with heterozygous females. Only female mice were used and they were 26-weeks old at the time of sacrifice. Two tissues i.e. cerebellum and skeletal muscle that are phenotypically affected due to *Sil1* mutations were used for proteomics, biochemical and morphological analyses.

3.2.1.4.2 Human

A skeletal muscle biopsy (Musculus quadriceps femoris) derived from an index patient (female, Caucasian) suffering from a phenotype similar to MSS, but not related to *SIL1* mutations was prepared for proteomics analysis at the Institute of Neuropathology, Aachen. For relative proteome comparison, the muscle lysates of two controls (matched for age, sex and muscle group) were also prepared at the same Institute.

3.2.2 Cell lysis and benzonase treatment

Human cells were lysed using 1% SDS buffer comprising 150 mM NaCl, 50 mM Tris-Cl, pH 7.8 and complete Mini (protease inhibitor cocktail). This procedure was carried out on ice and under a laminar flow hood. The volume of the lysis buffer was based on the assumption that 100 µL of buffer would be sufficient to solubilize e.g. one million cells. Typically, 200 - 300 µL of lysis buffer was used and cell homogenization was performed by pipetting the mixture up and down until the pellet was completely solubilized. In case of viscous cell lysates, benzonase (25

U/µL) was added with 2 mM $MgCl_2$ and the samples were incubated at 37°C for 30 min in order to degrade the nucleic acids (DNA and RNA) and reduce viscosity. Afterwards, the samples were clarified by centrifugation at 4°C and 18,000 rcf for 30 min. Finally, the clear supernatant was collected in a LoBind Eppendorf tube and stored at -80°C until further use.

3.2.3 Tissue processing

The brain section i.e. cerebellum and skeletal muscles derived from the same mouse i.e. either wild type or *woozy* (see Section 3.2.1.4.1) were processed/homogenized separately by mechanical grinding using mortar and pestle. In total, three biological replicates of each wild type and *woozy* were used in this work. Tissue dissection and grinding were performed on ice and under a laminar flow hood. After processing of each sample, all the equipment i.e. dissection platform, spatulas, mortar and pestle and ultrasonic probe were properly sterilized and dried completely before proceeding to the next sample.

3.2.3.1 Brain section

Approximately 1 mg of cerebellum obtained from six different mice was processed individually. The tissues were first snap frozen followed by mechanical grinding. Next, 600 µL of lysis buffer comprising 0.1% SDS, 1% sodium deoxychloate, 1% Triton X-100, 150 mM NaCl, 1mM $Na_2EDTA.2H_2O$, 50 mM Tris-Cl, pH 8.0 and complete Mini were added to the ground tissue and mixed to obtain a homogenous lysate. Next, the samples were treated with benzonase as described in section 3.2.2 and stored at -80°C until further use.

3.2.3.2 Skeletal muscle

The mammalian skeletal muscle fibers are extremely tough due to their inherent composition. Hence, a 2-step homogenization procedure was employed involving mechanical grinding and ultrasonication. First, ~1 mg of skeletal muscle was homogenized by grinding in a mortar using a pestle as described above. To the ground tissue, 600 µL of lysis buffer comprising 4% SDS, 10% 2-mercaptoethanol, 10% glycerol, 125 mM Tris-Cl, pH 6.4 and complete Mini were added and the grinding process was continued. Lysates were transferred into a LoBind Eppendorf tube and subjected to ultrasonication for 30 seconds (amplitude: 30; pulse 1s/1s) in the second step.

Next, the tissue lysates were clarified by centrifugation at 10°C and 25,000 rcf for 10 min. The supernatant was collected in a LoBind Eppendorf tube. To remove the harsh buffer components and make the samples amenable to the downstream sample preparation procedures (on-filter proteolysis, see Section 3.2.6), organic solvent protein precipitation was performed. An aliquot of 200 µL of tissue lysate was diluted 10-fold with ice cold EtOH in 1:10 ratio, briefly vortexed and stored at -40°C for 60 min followed by centrifugation in a pre-cooled (4°C) centrifuge at 18,000 rcf for 30 min. Next, the supernatant was discarded and the pellet was washed with 100 µL of ice cold acetone, briefly vortexed and centrifuged as mentioned above for 15 min. Finally, the supernatant was discarded and the protein pellet was dried under a laminar flow hood. Protein pellets were resolubilized with 6 M GuHCl, 50 mM NH_4HCO_3, pH 7.8 and stored at -80°C until further use.

3.2.4 Determination of protein concentration

Estimation of protein concentration of cell and tissue lysates was performed by a calorimetric bicinchoninic acid assay according to the manufacturer's instructions (Pierce BCA Protein Assay Kit). Briefly, three serial dilutions per sample were prepared with ultra-pure water and each dilution was aliquoted (25 µL) in triplicates into a 96-well plate. To each well, 200 µL of BCA solution (reagent A:B in 50:1 ratio) were added and the plate was incubated at 60°C for 30 min. Bovine serum albumin was used to determine the standard curve (5 - 250 µg/mL concentration range, five-point calibration, triplicate standards) and the absorbance was measured at a wavelength of 570 nm using a microtiterplate reader.

3.2.5 Carbamidomethylation

Cysteines were reduced with 10 mM DTT and incubation at 56°C for 30 min. The free thiol (-SH) groups were subsequently alkylated (carbamidomethylation) by the addition of 30 mM IAA and incubation at room temperature (RT) for 30 min in the dark.

3.2.6 Spin filter assisted sample preparation

Sample clean-up and proteolysis after carbamidomethylation were performed using molecular weight cut off (MWCO) membrane spin filters. This approach was first described by Manza et al.

[119], which was later modified and introduced as filter aided sample preparation (FASP) by Wiśniewski et al. [120]. This protocol was demonstrated to be effective in removing most of the lysis buffer components especially SDS, which is the major interfering contaminant in LC-MS analysis due to its strong ionization. The FASP protocol involves a centrifugal device consisting of a membrane filtration unit with nominal molecular weight cut-off (MWCO) of 10, 30 or 50 kDa. These MWCO sizes are based on an approximation for non-denatured proteins (mostly globular), but these spin filters are often used after (i) protein denaturation with SDS and urea as well as (ii) reduction and alkylation. Therefore, 30 kDa MWCO membrane spin filters were used in this work, which are also most commonly used by other groups in proteomics field [121]. Briefly, proteins in the sample are trapped in a high-MWCO filter unit whereas salts, low-molecular weight compounds and contaminants from samples flow through the spin filter and can be discarded. The use of urea buffer at high concentration (8.0 M) enables the removal of nearly all the surfactant (e.g. SDS) from the sample. Subsequent washes with an alkaline buffer e.g. 50 mM NH_4HCO_3, pH 7.8 eliminates urea from the spin filters. Finally, to the concentrated proteins, trypsin is added directly on to the spin filter for proteolytic digestion. After on-filter trypsin digestion, the tryptic peptides are eluted with the 50 mM NH_4HCO_3, pH 7.8 followed by a washing step with 0.5 M NaCl solution. The filter retains any high-molecular weight material such as partially digested or completely undigested proteins and debris. A description of FASP procedure performed in this work, albeit with slight changes using a 30 kDa MWCO filter is given below.

The sample lysate (after carbamidomethylation, see Section 3.2.5) was diluted with freshly prepared 8.0 M urea/100 mM Tris-Cl, pH 8.5 buffer [122] such that the concentration of urea was around 7.0 M. The lysate was placed on the spin filter and the device was centrifuged at 13,500 rcf at RT for 20 min. All the following centrifugation steps were performed at 13,500 rcf at RT for 15 min. To eliminate residual SDS, three washing steps were carried out with 100 µL of 8 M urea/100 mM Tris-HCl, pH 8.5. For buffer exchange, the spin filter was washed thrice with 100 µL of 50 mM NH_4HCO_3, pH 7.8. Here, the FASP workflow was slightly modified for instance 50 mM NH_4HCO_3, pH 7.8 was replaced with 50 mM triethylammonium bicarbonate (TEAB), pH 8.5 buffer as it does not contain primary amines as opposed to the former buffer to avoid desalting step prior to iTRAQ labeling. At first, for MSS fibroblasts and LCs cell lysates and their respective

controls, 50 mM NH_4HCO_3, pH 7.8 buffer was used, but it was changed later (as part of workflow optimization) to 50 mM TEAB, pH 8.5 buffer for processing mice tissue lysates. After buffer exchange, 100 µL of proteolysis buffer comprising: trypsin (1:40 w/w of enzyme to substrate ratio), NH_4HCO_3, pH 7.8 or TEAB pH 8.5, a chaotrope (GuHCl) and calcium chloride (as a source of Ca^{2+}) were added to the concentrated proteins and the spin filter was incubated at 37°C overnight. The latter two components (i.e. GuHCl and Ca^{2+}) were added to keep the proteins in a solubilized state and to stabilize trypsin [123], respectively. After incubation, the generated peptides were recovered by centrifugation followed by two consequent washing steps with 50 µL of 50 mM NH_4HCO_3, pH 7.8 (or 50 mM TEAB, pH 8.5). However, in the final step during peptides extraction, ultra-pure water was used instead of 0.5 M NaCl as mentioned in the original FASP protocol. The reason was to circumvent the desalting step in case 50 mM TEAB buffer, pH 8.5 was used. Test experiments that were conducted to compare the modified FASP protocol with the original one showed nearly identical results in terms of peptide recovery and subsequent protein identifications (Fig. 3.1 B).

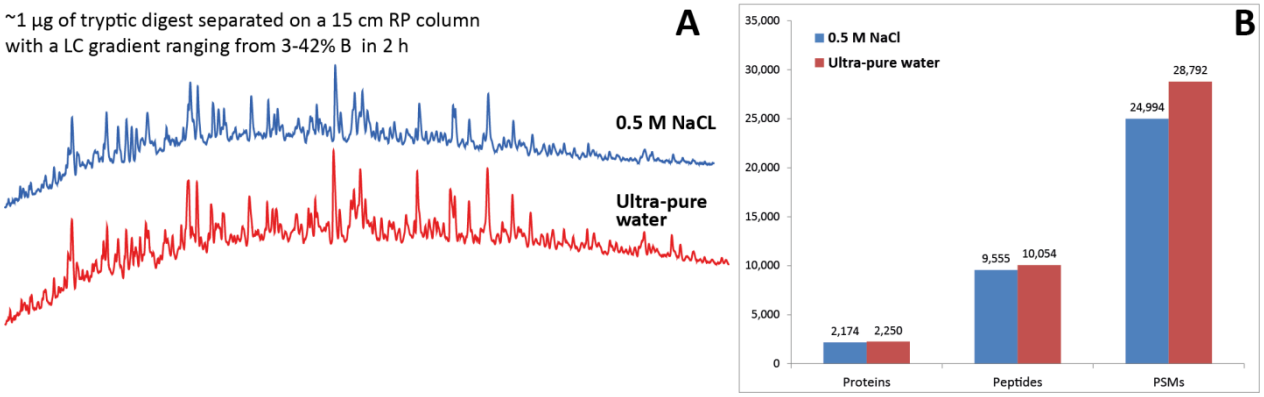


Figure 3.1: Human fibroblast cell lysates were processed with the FASP protocol and for peptides recovery (after overnight digestion with trypsin) in a second step, the spin filters were washed either with 0.5 M NaCl as in the original protocol or with ultra-pure water. LC-MS (see Section 3.2.13) and data analyses (see Section 3.2.14) were performed in the same manner for both conditions. (**A**) UV traces of the two conditions look nearly the same. (**B**) Database search results obtained from the raw MS data analysis of the two conditions show nearly same number of identifications in terms of proteins and peptides. However, the numbers of peptide spectrum matches (PSMs) were slightly higher (~13%) for ultra-pure water condition.

In case of mice skeletal muscle samples, the proteins that were solubilized with 6 M GuHCl buffer (see Section 3.2.3.2) were diluted 2-fold with 50 mM TEAB, pH 8.5 and placed on 30 kDa MWCO filters. Next, the washing steps and proteolysis were carried as described above. Thus generated tryptic peptides were acidified to a final pH < 3.0 with 10% TFA and stored at -80°C until further use.

3.2.7 In solution digestion of muscle lysates - human

The amount of protein in the skeletal muscle lysate obtained from an index patient and two controls was roughly estimated to be 5 µg per sample. Sample processing with FASP was not feasible since the components of the lysis buffer were not compatible with the MWCO membrane filters. For this reason, ethanol protein precipitation (see Section 3.2.3.2) followed by in solution digestion were performed. Next, protein pellets were resolubilized with 1 M GuHCl, 50 mM NH_4HCO_3, pH 7.8 and the samples were subjected to carbamidomethylation (see Section 3.2.5). Samples were then diluted with 50 mM NH_4HCO_3, pH 7.8 to a final GuHCl concentration of 0.2 M. Trypsin was added in 1:40 (w/w) of enzyme to substrate ratio and samples were incubated at 37°C overnight. After incubation, tryptic peptides were acidified to a final pH < 3.0 with 10% TFA and stored at -80°C until further use.

3.2.8 Evaluation of the digestion efficiency

Tryptic peptides were checked for digestion efficiency using a monolithic HPLC system as previously described [124]. Monolithic columns are more robust and sensitive than other techniques that are used for digestion control e.g. SDS-PAGE followed by coomassie or silver staining. Approximately, 1 µg of peptides were loaded on to a 200 µm x 5 mm pre-column with 0.1% TFA followed by separation on a 200 µm x 5 cm main column (both PepSwift, Thermo Scientific) with a 22 min LC gradient ranging from 10 - 50% of 84% ACN in 0.1% TFA at a flow rate of 2.2 µL/min using a UltiMate 3000 RSLC HPLC system.

3.2.9 Chemical labeling using iTRAQ reagents

In total four different iTRAQ-based experiments were performed in this work, which include one 4-plex and three 8-plex versions of the reagents. Whereas the 4-plex iTRAQ was used in the beginning, an iTRAQ 8-plex labeling procedure was established in the course of this work. The sample details, labeling scheme and the peptide amounts used for each label/condition are summarized below (Tables 3.6 - 3.8). Before labeling with iTRAQ tags, each sample (~1 µg) was analyzed on a nano-LC-MS system (see Section 3.2.13). The sample amounts were corrected based on the alignment of total ion chromatograms (TICs) to compensate for systematic errors

derived e.g. from the protein concentration estimation, such that each sample had identical starting material before labeling. In case of mice tissues (skeletal muscle and brain), where only three biological replicates were available, two pooled samples were generated by mixing equal amounts of peptides of respective biological triplicates (wild type or *woozy*).

Table 3.6: Labeling scheme of the human fibroblast cells with iTRAQ 4-plex.

Passage	Gender	Sample name	iTRAQ reagent	Peptide amount per label
9	Male	Healthy.1	114	100 µg
		Healthy.2	115	
		MSS2.1	116	
		MSS2.2	117	

Table 3.7: Labeling scheme of the human EBV-lymphoblastoid cells with iTRAQ 8-plex.

Passage	Gender	Sample name	iTRAQ reagent	Peptide amount per label
11	Male	Healthy.1	113	40 µg
		Healthy.2	114	
		Healthy.3	115	
		Healthy.4	116	
		MSS24	117	
		MSS32	118	
		MSS33	119	
		MSS94	121	

Table 3.8: Labeling scheme of mice cerebella and skeletal muscles with iTRAQ 8-plex.

Age	Gender	Sample name	iTRAQ reagent	Peptide amount per label
26 weeks	Female	Wildtype.1	113	30 µg
		Wildtype.2	114	
		Wildtype.3	115	
		Woozy.1	116	
		Woozy.2	117	
		Woozy.3	118	
		Wildtype mix	119	
		Woozy mix	121	

*numbers in the sample name column indicate biological replicate.

The iTRAQ labeling procedure with 4- or 8-plex reagents was carried out according to the manufacturer's instructions (AB SCIEX). The dried peptides were resolubilized with 0.5 M TEAB, pH 8.5, whereas the iTRAQ 4-plex reagents were diluted with ethanol and the 8-plex reagents with isopropanol (both LC-MS grade, 100%). In order to prevent rapid hydrolysis of the chemical tags, the organic concentration was kept ≥ 60% and 70% for 4 and 8-plex versions, respectively. Next, the individual label mixtures were added to the corresponding peptide sample solutions and were incubated at 25°C for 1 h (4-plex) or 2 h (8-plex). After incubation, the differentially

labeled samples were combined (multiplexing) and the reaction was quenched by adding 100 µL of ultra-pure water. The volume of the pooled peptide solution was reduced in a SpeedVac to ~20 µL, subsequently acidified with 10% TFA to a final pH < 3.0 and stored at -80°C until further use.

3.2.10 Desalting of proteolytic digests and iTRAQ labeled samples

Depending on the amount of peptides, the acidified (pH < 3.0) tryptic digests were desalted either with C18 Pipette Tips (10 µL and 100 µL; OMIX) or C18 solid phase extraction cartridges (SPEC, 4 mg sorbent, Agilent) using a vacuum manifold system. All volumes were appropriately used according to the column specifications i.e. size/capacity of the used tip/SPEC. First, the C18 material was activated with 100% ACN followed by equilibration with 0.1% TFA. Next, the peptide solutions were loaded on to the pre-equilibrated columns and the flow through was reloaded twice. The columns were then washed with 0.1% TFA and finally, the bound peptides were eluted with 60% ACN in 0.1% TFA. The eluates were completely dried in a SpeedVac and stored at -80°C until further use.

3.2.11 Off-line peptide fractionation

In this work, OFFGEL IEF and high-pH RP fractionation techniques were employed to reduce the sample complexity prior to LC-MS analysis. Both fractionation methods are powerful in terms of resolution, and a systematic comparative analysis between these two methods was performed together with BSc. Jennifer Baumann (Bachelor thesis under my supervision). This showed better orthogonality for the high-pH RP approach. Furthermore, the high-pH RP fractions could be directly subjected to LC-MS analysis after solvent evaporation thus avoiding the laborious desalting procedure required for OFFGEL fractions, which might also lead to sample losses.

3.2.11.1 OFFGEL isoelectric focusing

The dried iTRAQ 4-plex labeled and desalted peptides (human fibroblasts) were separated using an OFFGEL Fractionator with a 24-well set-up. The peptide samples were resolubilized to a final volume of 3.6 mL using the OFFGEL peptide stock solution (12% glycerol plus 1.2% carrier ampholytes). Prior focusing, the IPG gel strip with a linear pH range 3 - 10 was treated with the

peptide IPG strip rehydration solution (OFFGEL peptide stock plus 0.2% water) for 15 min. Next, 150 µL of sample solution was loaded in each well. The IEF of the peptides was performed at 50 µA starting current and a maximum power supply of 200 mW. The voltage was increased from 300 to 8,000 V until 60 kVh were reached. After focusing, the separated peptide fractions were immediately transferred into LoBind Eppendorf tubes. Each individual fraction was acidified with 10% TFA to a final pH < 3.0 and desalted as described in section 3.2.10. The eluates were completely dried in a SpeedVac and resolubilized in 0.1% TFA prior to nano-LC-MS analysis.

3.2.11.2 High-pH RP fractionation

In general, the RP fractionations can be conducted at pH 10.0 [125] albeit, high pH conditions can affect column and capillary stability due to the hydrolysis of siloxane groups [126]. Therefore, in this work all RP fractionations were performed at pH 6.0 using an UltiMate 3000 HPLC system. The desalted and dried peptides were resolubilized in buffer A (10 mM ammonium acetate, 0.4 mM FA, pH 6.0) and fractionated on a C18 RP column (Table 3.9) with a binary buffer system; buffer A: 10 mM ammonium acetate, 0.4 mM FA, pH 6.0 and buffer B: 84% ACN in 10 mM ammonium acetate, 0.4 mM FA, pH 6.0. Peptides were loaded onto the column with buffer A at a flow rate of 12.5 µL/min and separation was carried out using the following gradient: 3% B for 10 min, 3-50% B in 65 min, 50-60% B in 5 min, 60-95% B in 5 min, 95% B hold for 5 min, 95%-3% B in 5 min and finally column re-equilibration with 3% B for 20 min.

Table 3.9: Details of the offline RP fractionation carried out at pH 6.0 using an UltiMate 3000 HPLC.

Sample details	RP Column details	Fraction collection details
iTRAQ 8-plex labeled human EBV-LCs; 50 µg multiplexed sample	ZORBAX 300SB C18, 0.5 mm ID, 15 cm length, 5 µm particle size, 300 Å pore size	24 fractions were collected at 30 sec intervals from min 15 to 85 in concatenation mode
iTRAQ 8-plex labeled mice cerebella; 50 µg multiplexed sample		
iTRAQ 8-plex labeled mouse skeletal muscles; 25 µg multiplexed sample	BioBasic C18, 0.5 mm ID, 15 cm length, 5 µm particle size, 300 Å pore size	24 fractions were collected at 60 sec intervals from min 15 to 75 in concatenation mode
Unlabeled human RCMH (myoblastic cell line); 25 µg tryptic digest		16 fractions were collected at 60 sec intervals from min 15 to 75 in concatenation mode

Each fraction was collected at 30s (or 60 s) time interval and in total 24 (or 16) fractions were collected in a LoBind Eppendorf tubes from 15 min to 75 min (or 85 min) retention time . After

collecting the first 24 (or 16) fractions, the collection process was repeated from the first collection tube i.e. in a concatenation mode to improve the orthogonality of the analysis. After fractionation, all peptide fractions were completely dried in a SpeedVac and resolubilized in 0.1% TFA prior to nano-LC-MS analysis.

3.2.12 *In vitro* carbamylation

To optimize sample preparation methods for LC-MS analysis, the occurrence of protein carbamylation which can be artificially induced through urea was evaluated in this work. Due to its strong chaotropic nature, urea is the most commonly used denaturing agent in the proteomics workflows and is an essential part of the here used FASP protocol for sample preparation and digestion. However, a possible drawback of using urea is the potentially occurring *in vitro* carbamylation of primary amines (occurring due to decomposition of urea to isocyanate, over time and preferably upon heating), which results in an undesired artefact, especially hampering quantitative studies that are based on labeling of primary amines.

3.2.12.1 Peptide stock mixture and *in vitro* carbamylation conditions

Protein stock solutions of β-casein, β-lactoglobulin, hemoglobin, serum albumin (all bovine) and myoglobin (horse) were prepared in ultra-pure water at a concentration of 1 mg/mL. 100 µg of each protein solution was diluted 10-fold with 50 mM NH_4HCO_3, pH 7.8. All five proteins were processed independently. Carbamidomethylation was performed as described in section 3.2.5. Trypsin was added in 1:50 (w/w) of enzyme to substrate ratio and proteins were digested at 37°C overnight. The generated tryptic peptides of each protein were acidified with 10% TFA to a final pH < 3.0 followed by SPE (see Section 3.2.10). To prepare the stock peptide mixture, the eluates corresponding to 800 pmol of each digested protein were pooled. This combined peptide mixture was then divided into 50 pmol aliquots, which were completely dried in a SpeedVac. Next, each dried peptide mixture aliquot was resolubilized in 50 µL [final concentration: 1 pmol/µL] of freshly prepared urea buffers (see below). The peptide mixture was then used to evaluate the extent of urea-induced carbamylation under below mentioned conditions typically used in bottom-up proteomics workflows and each experimental setup was performed in triplicate. Moreover, to find if degradation of urea into isocyanate occurs in solid

form over long term storage, each incubation condition was performed once with ~4 month old and once with ~8 year old urea. The conditions were: (1) control (in 0.1% TFA), (2) harsh treatment (8.0 M urea, 100 mM Tris-HCl, pH 8.5; 61°C, 15 h), (3) reduction of disulfide bonds (8.0 M urea, 100 mM Tris-HCl, pH 8.5; 56°C, 30 min) [127], different overnight digestion conditions with trypsin at 37°C for 15 h, i.e. (4) 2.0 M urea, 50 mM NH_4HCO_3, pH 7.8 [128], (5) 2.0 M urea, 100 mM Tris-HCl, pH 8.5 [129], (6) 0.1 M urea, 50 mM NH_4HCO_3, pH 7.8, (7) overnight digestion in presence of GuHCl instead of urea (0.2 M GuHCl, 50 mM NH_4HCO_3, pH 7.8; 37°C, 15 h), (8) overnight digestion with Lys-C, a serine protease that cleaves at the C-terminus of Lys residues (8.0 M urea, 100 mM Tris-HCl, pH 8.5; 20°C, 18 h) [120].

3.2.12.2 Two-step digest of fibroblast cells

Next, the extent of urea-induced *in vitro* carbamylation occurring during preparation of complex samples e.g. as a total cell lysate was evaluated. Approximately, 250 µg of fibroblast cell pellets were lysed with 100 µL of RapiGest SF (surfactant) solution i.e. lyophilized RapiGest powder was reconstituted in 50 mM NH_4HCO_3, pH 7.8 to get a final concentration of 0.1% w/v. Benzonase and protein concentration determination steps were performed as described in sections 3.2.2 and 3.2.4, respectively. 20 µg protein aliquots for each urea-based and urea-free (used as a positive control) processing were taken. In case of urea-based, samples were diluted 10-fold with 8.0 M urea/100 mM Tris-HCl, pH 8.5, whereas urea-free samples were diluted 10-fold with RapiGest solution. Carbamidomethylation was performed as previously described (see Section 3.2.5), but with 5 mM DTT and 15 mM IAA, respectively. In the first step, Lys-C was added in 1:100 (w/w) of enzyme to substrate ratio and the samples were incubated at 20°C for 18 h. For the second step, the urea concentration was lowered from 8.0 M to 2.0 M with 100 mM Tris-HCl, pH 8.5 in both sample sets. Proteolysis was continued with trypsin (1:100 w/w) at 20°C for 4 h. Next, the digests were acidified with 10% TFA to a final concentration of 0.4% and incubated at 37°C for 45 min for the hydrolysis of RapiGest. After incubation, the samples were centrifuged at 13,000 rcf for 10 min to pellet the hydrolytic byproducts of the surfactant and the clear supernatant was collected for nano-LC-MS analysis.

3.2.13 Nano-LC-ESI-MS analysis

All peptide separations were carried out on C18 RP columns with a binary gradient (buffer A: 0.1% FA; buffer B: 84% ACN in 0.1% FA, pH 2.7) using nano-flow U3000 HPLC or U3000 RSLC HPLC systems (Thermo Scientific). The C18 RP columns used were either commercial or self-packed (in house). The dimensions of the commercial C18 Acclaim PepMap (Thermo Scientific) column were: trapping column 100 µm inner diameter x 2 cm length, 5 µm particle size, 100 Å pore size, and main column 75 µm inner diameter x 15 cm or 50 cm length, 2 µm particle size, 100 Å pore size. The self-packed columns were filled with Kinetex C18 material (Phenomenex) having 2.6 µm particle size, 100 Å pore size and their dimensions were: trapping column 100 µm x 2 cm and main column 75 µm x 30 cm. All peptide solutions were prepared in 15 µL of loading buffer i.e. 0.1% TFA prior to LC-MS analysis. Peptides were first preconcentrated on the trapping column using 0.1% TFA followed by separation on the main column using the above mentioned binary gradient ranging from 3-42% B for 50 min, 120 min, 127 min, 200 min at a flowrate of 230 nL/min, 250 nL/min.

The following mass spectrometers were employed for performing MS and MS/MS analysis: LTQ Orbitrap XL, LTQ Orbitrap Velos, Orbitrap Elite and Q Exactive (all from Thermo Scientific). ESI was used as an interface between the HPLC and MS. All MS were operated in a DDA mode. The key DDA parameter settings, such as the number of MS/MS scan events (Top N), dynamic exclusion duration, automatic gain control (AGC) target values i.e. the number of charges to inject and analyze and maximum ion injection times (IT) i.e. the maximum time that ions are allowed to accumulate in the linear ion trap or the C-trap, before being transferred to the Orbitrap were selected according to the complexity of the sample and performance characteristics of the respective MS. The type of fragmentation for MS/MS was chosen depending on the purpose of the analysis e.g. HCD for iTRAQ labeled samples. Furthermore, internal calibration of the Orbitrap was done by using the polysiloxane ion at m/z 371.101236 as lock mass [130] on all MS instruments. During MS analysis of the iTRAQ-labeled samples, a 10% (v/v) NH_4OH solution was placed in front of the ESI source for charge state reduction since isobaric tags increase the peptide charge states during ESI-MS consequently decreasing protein identification rates [131]. As different LC-MS combinations have been used during this work,

instead of listing all the parameters for each sample set, some examples will be given for the respective sample type. An overview of the LC-MS instruments and the parameters used for data acquisition for individual experiments is given in the Appendices 10.1 and 10.2. The HPLC and MS instruments were controlled by Chromeleon and Xcalibur software (both Thermo Scientific), respectively.

MSS fibroblasts - iTRAQ 4-plex

Each OFFGEL fraction (total 24 samples) was analyzed using an Ultimate 3000 HPLC system coupled to an LTQ-Orbitrap Velos. Peptide solutions were preconcentrated on a trapping column for 20 min using 0.1% TFA at a flow rate of 5 µL/min followed by separation on a main column (both self-packed) with a 200 min LC gradient ranging from 3-45% B at a flow rate of 230 nL/min. Full MS scans were acquired in the Orbitrap from m/z 300 to 2,000 at a resolution of 60,000 (at m/z 400) after accumulation of 1 x 10^6 ions (AGC target value) with a maximum IT of 50 ms. MS/MS events were triggered from the full survey scan and based on the minimum intensity threshold of 5,000. Dynamic exclusion duration for 12 s was applied (at a repeat count of 1, exclusion list size of 500) with a 5 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on and only charge states from +2 to +4 were selected for fragmentation. The five most intense precursor ions were subjected to MS/MS in a HCD collision cell with an AGC value of 5 x 10^3 ions and a maximum IT of 100 ms with an isolation window of ±1.5 m/z. Precursor ion activation was done with an activation time of 0.1 ms and a normalized collision energy (NCE) of 45% was used. Slightly higher collision energy is required for iTRAQ-labeled peptides owing to the presence of isobaric tags to increase the fragmentation efficiency. HCD is preferred over CID fragmentation (in the LTQ) as the former method has no low-mass cutoff and the iTRAQ reporter ions (typically observed from m/z 113) can be extracted and subsequently detected. The MS/MS scans were acquired in the Orbitrap at a resolution of 7,500 (at m/z 400) with a starting mass of 105 m/z.

Mice skeletal muscles - iTRAQ 8-plex

Each pH 6.0 fraction (total 24 samples) was analyzed using an Ultimate 3000 RSLC HPLC system coupled to a Q Exactive (see Fig. 1.13). Peptides were preconcentrated on a trapping column for

10 min using 0.1% TFA at a flow rate of 20 µL/min followed by separation on a 50 cm main column (both Acclaim PepMap) with a 127 min LC gradient ranging from 3-42% B at a flow rate of 250 nL/min. The full MS scans were acquired from m/z 200 to 2,000 at a resolution of 70,000 (at m/z 200) with an AGC target value of 1×10^6 ions and maximum IT of 120 ms. Isolation of precursors was performed by the quadrupole with a window of 2.0 m/z. The fifteen most intense ions were fragmented in the HCD cell with an AGC target value of 2×10^5 ions and maximum IT of 120 ms with a NCE of 30%. MS/MS scans were acquired at a resolution of 17,500 with a starting mass of 105 m/z taking into account a dynamic exclusion of 20 s. Precursor ions with charge states of +1, > +5 or unassigned were excluded from MS/MS analysis. The "underfill" ratio, which specifies the minimum percentage of the target value likely to be reached at maximum fill time, was defined as 10%, which corresponds to a minimum precursor intensity of 1.7×10^5 to trigger a MS/MS scan.

SIL1-depleted HEK293 cells - Label-free analysis

In total two sample sets, each comprising one control (Scr) and two SIL1-depleted cell lines (ΔSIL1_1 and ΔSIL1_2) were processed and each condition was measured in triplicate, resulting in a total of 18 LC-MS runs. Each condition (1 µg each) was analyzed using an Ultimate 3000 RSLC coupled to an Orbitrap Elite (see Fig. 1.12). Peptides were separated as described above using a 187 min LC gradient ranging from 3-42% B at a flow rate of 250 nL/min. Full MS survey scans were acquired in the Orbitrap from m/z 300 to 1,500 at a resolution of 60,000 with an AGC target value of 1×10^6 ions and maximum IT of 100 ms. MS/MS events were triggered from the full scan and with a minimum intensity threshold of 2,000. A dynamic exclusion of 30 s was used (at a repeat count of 1, exclusion list size of 500) with 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on and only charge states from +2 to +4 were selected for fragmentation. Fragmentation of the 15 most intense signals were subjected to ion trap CID with an AGC value of 1×10^4 ions and a maximum IT of 100 ms at an isolation window of ±2.0 m/z. CID spectra were generated with a NCE of 35% and an activation time of 10 ms.

SIL1-depleted HEK293 cells - Targeted-MS analysis

For targeted MS analysis, 27 proteins that are known to be involved in the ER stress and the UPR pathway [22b, 132] were selected and relatively quantified between control and SIL1-depleted HEK293 cells using an in house developed PRM-based assay on Q Exactive instrument (see Section 1.4.2.3). The assay method development was done using Skyline software [117] as follows. At first, the peptides were selected from *in silico* digested FASTA sequences of the respective proteins using the human UniProt database (see Table 3.11) as a background proteome. Next, these peptides were filtered according to the following criteria for ensuring reliable protein quantitation: (i) uniqueness, (ii) fully tryptic with no missed cleavage sites, (iii) a length of 8-25 amino acid residues and (iv) no Met residues. After filtering, a total of 76 precursors with charge states +2 or +3 were exported as inclusion list from Skyline for an unscheduled (i.e. no retention time boundaries for the precursor m/z were specified) targeted PRM analysis. Simultaneously, a Q Exactive spectral library was generated using the identified peptide MS/MS data of all these proteins obtained from conducted label-free DDA analyses to manually validate the peak assignments of the acquired PRM data.

Each condition was analyzed in triplicate (9 samples, 1 µg each) and peptides were separated using an Ultimate 3000 RSLC system as described above with a 127 min LC gradient ranging from 3-42% B at a flow rate of 250 nL/min. The Q Exactive was operated in a PRM mode and precursor isolation was performed by the quadrupole in the front end with an isolation width of 2.0 m/z. Fragmentation was performed in the HCD cell with a NCE of 27% and up to 18 PRM scans were acquired at a resolution of 70,000 with an AGC target value of 1×10^6 ions and a maximum injection time of 150 ms as triggered by the scheduled inclusion list. Thus generated raw MS/MS data were imported into Skyline for visualization, manual validation and assignment of retention time boundaries (predicted by Skyline). Next, the analysis of all 9 samples (1 µg each) was repeated with same LC-MS setup as described above, but with defined retention time windows information for each individual precursor m/z (in total 76) was added in the inclusion list. To avoid the loss of precursor peptides (most likely due to unpredicted retention time shifts during LC separation), the retention time boundaries were increased by ±3.5 min around the predicted value from the Skyline (i.e. 2 min).

3.2.14 Database searches

Database searches and data analysis of iTRAQ samples were performed using Proteome Discoverer or PD (Thermo Scientific). Whereas, Progenesis from Nonlinear Dynamics (Newcastle upon Tyne, U.K.), searchGUI [133] and PeptideShaker software [134] were used to process label-free data. To maximize the number of PSMs, a multiple-search engine strategy was applied using the same set of search parameters. In PD, the raw MS data can be directly imported and all steps in the protein identification/quantification pipeline can be performed in an automated fashion based on a user-defined workflow. PD can generate peak lists from the raw data files, perform sequence database searches with single or multiple algorithms, combines the output in case a multiple-engine search strategy is applied, validates protein identifications using a FDR on the PSM level (typically 1% using Peptide Validator or Percolator settings) and finally, combines this information with quantitative information per peptide and protein (e.g. iTRAQ). SearchGUI, however, only supports mascot generic format (mgf) files as the input format for the MS/MS searches and ProteoWizard [135] was used to convert raw data to mgf. Similar to PD, searchGUI also facilitates multiple-search engine strategy, albeit requires a platform to combine, visualize and analyze the identification results. Here, PeptideShaker software, which is a search engine independent platform, was used for the interpretation of search results from multiple engines and also for validating the data using a FDR (1%) on the PSM, peptide and protein level.

In both PD and searchGUI, the following parameters were selected for the database searches of all experiments. Trypsin was selected as protease allowing a maximum of two missed cleavages and the searches were performed using Mascot [90], Sequest [89], X! Tandem [136], MS-GF+ [137] algorithms in a target/decoy mode against either a human or a mouse UniProt database (both downloaded on 11th of December 2013, containing 20,273 and 16,649 target entries, respectively. see Table 3.10). Whereas, the database searches of the peptide mixtures from the carbamylation experiment were performed against a merged database containing the amino acid sequences of the five model proteins taken from UniProt, in a yeast background taken from *Saccharomyces* genome database summing up to a total 6,723 target sequences to facilitate FDR calculation. Carbamidomethylation of Cys and oxidation of Met were set as fixed and variable modifications, respectively. Precursor ion mass tolerance was set to 10 ppm (MS scans

acquired in the Orbitrap), whereas product ion tolerances were set to 0.5 Da and 0.02 Da for MS/MS data acquired in the linear ion trap and in the Orbitrap, respectively. In case of iTRAQ experiments, iTRAQ-related modifications i.e. 4 and 8-plex versions were set on N-terminus and Lys as fixed modifications and the vendor (AB SCIEX) provided isotope purity correction factors were implied in the reporter ions quantifier node of PD software for the iTRAQ data analysis (Table 3.11).

Table 3.10: Details of the protein sequence databases. The decoys were generated by reversing the forward target sequences.

Name	Download date	Organism	Target Sequences	Source
UniProt database	30.07.2012	Human	20,232	http://www.uniprot.org
UniProt database	11.12.2013	Human	20,273	http://www.uniprot.org
UniProt database	11.12.2013	Mouse	16,649	http://www.uniprot.org

Table 3.11: Used algorithms and software for database searches. Cys carbamidomethylation and Met oxidation were used as fixed and variable modifications, respectively for all the datasets.

Experiment name	Search algorithms	Software	Protein database	Additional modifications		Mass tolerances	
				Variable	Fixed	MS	MS/MS
Carbamylation: peptide mixtures	Mascot	PD	Merge SGD, 08.05.2012	N-term, K, R (Carbamyl)		10 ppm	0.5 Da
Carbamylation: fibroblast 2-step digest	Mascot, SEQUEST	PD	Human, 30.07.2012	N-term, K, R (Carbamyl)		10 ppm	0.5 Da
Human fibroblasts: iTRAQ 4-plex	Mascot, SEQUEST	PD	Human, 30.07.2012		N-term, K (iTRAQ 4-plex)	10 ppm	0.02 Da
Human EBV-lymphoblasts: iTRAQ 8-plex	Mascot, SEQUEST	PD	Human, 16.09.2014		N-term, K (iTRAQ 8-plex)	10 ppm	0.02 Da
SIL1-depleted HEK293: label free	Mascot, OMSSA, X! Tandem	SearchGUI	Human, 11.12.2013			10 ppm	0.5 Da
Human RCMH: label free	Mascot, MS-GF+, X! Tandem	SearchGUI	Human, 11.12.2013			10 ppm	0.5 Da
Mice cerebella and cerebra: iTRAQ 8-plex	Mascot, SEQUEST	PD	Mouse, 11.12.2013		N-term, K (iTRAQ 8-plex)	10 ppm	0.02 Da
Mice skeletal muscles: iTRAQ 8-plex	Mascot, SEQUEST	PD	Mouse, 11.12.2013		N-term, K (iTRAQ 8-plex)	10 ppm	0.02 Da
Human skeletal muscles: label free	Mascot, X! Tandem	SearchGUI	Human, 16.09.2014			10 ppm	0.02 Da

3.2.15 Data analysis and statistical evaluation

Data analyses including statistical evaluation were carried out as described below using Microsoft Excel 2010.

3.2.15.1 iTRAQ data

The results from PD were exported with the following data filtering criteria: PSMs with FDR < 1%, search engine rank 1 and proteins that were quantified with ≥ 2 unique peptides.

MSS fibroblasts: iTRAQ 4-plex data

To normalize the iTRAQ ratio data, which is inherently asymmetrical (i.e. skewed), the ratio data was transformed into log values to a base of 2. Values with a log2-ratio of zero represent equal expression; values of -1 and 1 represent twofold down- and twofold upregulation, respectively. As a general assumption, in most cases, only a relatively low subset of proteins is expected to be

differentially regulated and therefore the majority of the log-ratios are expected to be centered around zero. However, due to the experimental bias e.g. pipetting errors or deviating BCA assay results, the median of the log-transformed ratios might differ from zero (Fig. 3.2 MD1). In such cases, the individual channel log2-transformed ratios of each protein were zero-centered by subtracting the protein specific medians taken across all the samples to compensate for the systematic errors. In PD, four different ratios were generated by using two replicates of MSS2 patient and healthy controls i.e. MSS2.1/Healthy.1, MSS2.1/Healthy.2, MSS2.2/Healthy.1 and MSS2.2/Healthy.2. Firstly, the individual channel ratios of each protein were log2-transformed and for each channel an overall median (MD1) of all proteins was calculated (Fig. 3.2).

A	B	C	D	E	F	G	H	I	J	K
			Median over all proteins per channel (**MD1**)				MSS2.1/ Healthy.1	MSS2.1/ Healthy.2	MSS2.2/ Healthy.1	MSS2.2/ Healthy.2
							-0.57	-0.50	-0.53	-0.46
UniProt Accession	Protein	Gene	116/114	116/115	117/114	117/115	116/114	116/115	117/114	117/115
P55290	Cadherin-13	CDH13	2.03	2.15	2.15	2.28	1.02	1.10	1.10	1.19
Q99439	Calponin-2	CNN2	1.59	1.69	1.67	1.74	0.67	0.76	0.74	0.80
Q5KU26	Collectin-12	COLEC12	0.31	0.32	0.33	0.34	-1.71	-1.66	-1.62	-1.57
P07585	Decorin	DCN	1.86	2.02	1.36	1.59	0.90	1.02	0.45	0.67
			Raw iTRAQ ratios				Log2-transformed ratios			

Figure 3.2: Four ratios were obtained from PD (columns **D** - **G**) and each individual channel ratio of each protein was log2-transformed (columns **H** - **K**).

Secondly, each channel MD1 was subtracted from log2-transformed ratio of each protein to get normalized ratios, which were then averaged to get a final MSS2/healthy ratio. Thirdly, a global median ratio (MDglobal) and standard deviation (SDglobal) were calculated by considering all corresponding protein ratios (MSS2/Healthy). Lastly, relative standard deviations (RSD) were calculated for each protein by using the normalized ratios of the four individual ratios (Fig. 3.3).

A	B	C	L	M	N	O	P	Q	R	S	T	U	V
			MSS2.1/ Healthy.1	MSS2.1/ Healthy.2	MSS2.2/ Healthy.1	MSS2.2/ Healthy.2	0.51	SDglobal	MSS2.1/ Healthy.1	MSS2.1/ Healthy.2	MSS2.2/ Healthy.1	MSS2.2/ Healthy.2	
	Zero-centered median per channel		0.00	0.00	0.00	0.00	0.00	MDglobal	1.00	1.00	1.00	1.00	
UniProt Accession	Protein	Gene	116/114	116/115	117/114	117/115	AVG ratio (log2)	MSS2/ Healthy	116/114	116/115	117/114	117/115	RSD%
P55290	Cadherin-13	CDH13	1.59	1.60	1.64	1.64	1.62	3.07	3.01	3.04	3.11	3.12	2%
Q99439	Calponin-2	CNN2	1.24	1.25	1.27	1.25	1.25	2.39	2.36	2.39	2.41	2.38	1%
Q5KU26	Collectin-12	COLEC12	-1.14	-1.16	-1.08	-1.11	-1.13	0.46	0.45	0.45	0.47	0.46	2%
P07585	Decorin	DCN	1.47	1.51	0.98	1.12	1.27	2.41	2.77	2.86	1.97	2.18	18%
			Normalized log2-ratios after median subtraction					Final ratios	Back-transformed normalized log2-ratios for calculating RSD				

Figure 3.3: The overall median (MD1) of the respective channel was subtracted from the log2-transformed ratio of each protein (columns **L** - **O**) such that the overall median of the log2-ratios per channel is now centered around zero. An average (AVG) ratio of MSS/Healthy was calculated using the normalized ratios across all channels for each protein (column **Q**). Next, MDglobal and SDglobal were calculated using MSS/Healthy over all proteins. To estimate the variation between the replicates, relative standard deviation (RSD) values (column **V**) were calculated using the back-transformed log-2 normalized ratios of each protein (columns **R** - **U**).

Only proteins with RSD values ≤ 20% and with log2-ratios ≥ 2*SDglobal away from the MDglobal were considered as regulated.

MSS-LCs and mice tissues (skeletal muscle and brain): iTRAQ 8-plex data

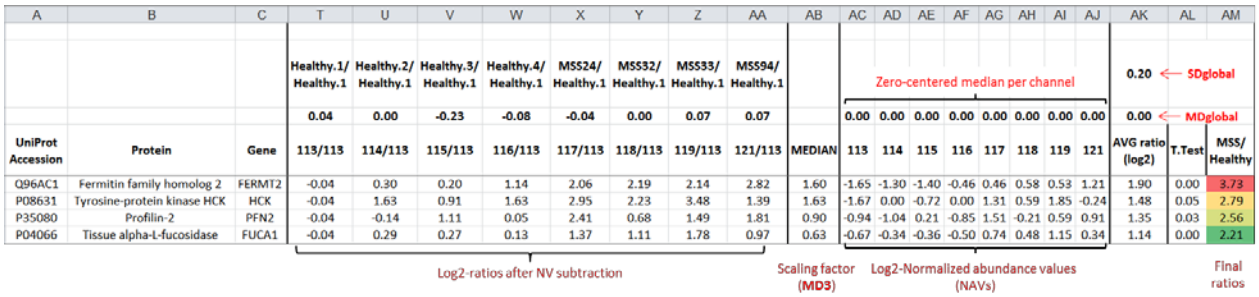
As the iTRAQ 8-plex data comprised 3 - 4 biological replicates per condition, instead of working with 7 ratios against one reference sample, as provided by PD, here ratios were transformed into normalized abundance values (NAVs). Thus, for statistical comparison e.g. Student's T-Test, a hypothetical ratio (i.e. 113/113) was generated in order to have eight data points in an experiment. The individual channel ratios of each protein were log2-transformed and for each channel MD1 over all proteins was calculated (same as above). Next, a second median (MD2) was generated by taking the MD1s across all channels. MD2 was then subtracted from the individual MD1s to deduce the normalization factors (NFs) for each channel (Fig. 3.4).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
											-0.04 ← MD2							
											Normalization factor (NF) per channel →	0.04	0.00	-0.23	-0.08	-0.04	0.00	0.07	0.07
											Median (MD1) over all proteins per channel →	0.00	-0.04	-0.27	-0.12	-0.08	-0.05	0.02	0.02
UniProt Accession	Protein	Gene	113/113	114/113	115/113	116/113	117/113	118/113	119/113	121/113	113/113	114/113	115/113	116/113	117/113	118/113	119/113	121/113
Q96AC1	Fermitin family homolog 2	FERMT2	1	1.24	0.98	2.09	4.06	4.54	4.61	7.39	0.00	0.31	-0.03	1.06	2.02	2.18	2.20	2.88
P08631	Tyrosine-protein kinase HCK	HCK	1	3.10	1.60	2.94	7.49	4.66	11.67	2.75	0.00	1.63	0.68	1.55	2.90	2.22	3.54	1.46
P35080	Profilin-2	PFN2	1	0.91	1.85	0.98	5.17	1.60	2.94	3.67	0.00	-0.13	0.88	-0.03	2.37	0.68	1.56	1.88
P04066	Tissue alpha-L-fucosidase	FUCA1	1	1.23	1.03	1.04	2.52	2.15	3.60	2.05	0.00	0.30	0.04	0.05	1.33	1.11	1.85	1.03
			Raw iTRAQ ratios								Log2-transformed ratios							

Figure 3.4: A hypothetical ratio i.e. 113/113 was created (column **D**). Four ratios were obtained from PD and each individual channel (columns **D** - **K**) of each protein was log2-transformed (columns **L** - **S**).

These NFs were then used to normalize the respective log2-transformed ratios for each protein per channel to obtain normalized ratios (NR). Next, a third median (MD3 or scaling factor, Fig. 3.5, column AB) was calculated by taking the NRs for each protein across all channels. This MD3 was then subtracted from the NR of each protein to get NAVs that were then grouped accordingly (4 healthy or 4 MSS patients) for each protein. Next, the Student's T-Test p-values with a significance level of ≤ 0.05 (two-sample assuming unequal variance) and ratios (i.e. MSS/Healthy) were calculated. Finally, MDglobal and SDglobal were calculated by considering all corresponding protein ratios i.e. MSS/healthy. Only proteins with Student's T-Test p-values ≤ 0.05 and with log2-ratios ≥ 2*SDglobal away from the MDglobal were considered as potentially regulated. In case of mice skeletal muscle and brain tissue samples, the pooled sample of three

biological replicates of wild type or *woozy* was grouped with the respective condition to have four data points for each biological state (Fig. 3.5).



A	B	C	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
			Healthy.1/ Healthy.1	Healthy.2/ Healthy.1	Healthy.3/ Healthy.1	Healthy.4/ Healthy.1	MSS24/ Healthy.1	MSS32/ Healthy.1	MSS33/ Healthy.1	MSS94/ Healthy.1		Zero-centered median per channel								0.20 ← SDglobal		
			0.04	0.00	-0.23	-0.08	-0.04	0.00	0.07	0.07		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00 ← MDglobal		
UniProt Accession	Protein	Gene	113/113	114/113	115/113	116/113	117/113	118/113	119/113	121/113	MEDIAN	113	114	115	116	117	118	119	121	AVG ratio (log2)	T.Test	MSS/ Healthy
Q96AC1	Fermitin family homolog 2	FERMT2	-0.04	0.30	0.20	1.14	2.06	2.19	2.14	2.82	1.60	-1.65	-1.30	-1.40	-0.46	0.46	0.58	0.53	1.21	1.90	0.00	3.73
P08631	Tyrosine-protein kinase HCK	HCK	-0.04	1.63	0.91	1.63	2.95	2.23	3.48	1.39	1.63	-1.67	0.00	-0.72	0.00	1.31	0.59	1.85	-0.24	1.48	0.05	2.79
P35080	Profilin-2	PFN2	-0.04	-0.14	1.11	0.05	2.41	0.68	1.49	1.81	0.90	-0.94	-1.04	0.21	-0.85	1.51	-0.21	0.59	0.91	1.35	0.03	2.56
P04066	Tissue alpha-L-fucosidase	FUCA1	-0.04	0.29	0.27	0.13	1.37	1.11	1.78	0.97	0.63	-0.67	-0.34	-0.36	-0.50	0.74	0.48	1.15	0.34	1.14	0.00	2.21

Log2-ratios after NV subtraction | Scaling factor (MD3) | Log2-Normalized abundance values (NAVs) | Final ratios

Figure 3.5: The median subtracted log2-ratios (columns **T** - **AA**) of each protein across all channels were used to calculate a median or scaling factor (column **AB**) to zero-center the medians. After subtracting the scaling factor from each normalized ratio, the generated normalized abundance values (columns **AC** - **AJ**), were grouped accordingly and an average (AVG) ratio of MSS/Healthy was calculated for each protein (column **AK**). Next, MDglobal and SDglobal were calculated using MSS/Healthy ratios over all proteins. To estimate the significance between the conditions, Student's T-Test p-values (with a significance level of 0.05) were generated (column **AL**). The final ratios (column **AM**) were calculated by back-transforming log-2 average ratio (AVG, column **AK**) of each protein.

3.2.15.2 Label-free data

3.2.15.2.1 Precursor area quantification

Quantitative analysis of the acquired label-free MS data of SIL1-depleted HEK293 cells was performed using the Progenesis software. The triplicate measurements of both sample sets i.e. ΔSIL1_1 and ΔSIL1_2, respectively, were compared to the corresponding control triplicates (Scr) separately. After importing the MS raw files, data processing including (i) feature/peptide extraction, (ii) selection of the reference LC-MS run, (iii) alignment, (iv) peak picking and (v) normalization was done automatically by Progenesis. A feature is defined as the sum of all the MS signals produced by the same peptide ion across all the samples analyzed and has specific characteristics such as retention time, m/z, charge and intensity [138]. Next, the features within a set of defined parameters i.e. retention time and m/z windows with charge states from +2 to +4 were considered for peptide statistics and ANOVA. Furthermore, given the nature of label-free quantification, where many samples are measured under the exact same conditions and therefore produce a considerable extent of redundant (or overlapping) MS/MS information for a single feature across multiple LC-MS runs, the number of redundant MS/MS spectra used for database search was reduced. Therefore, for a given feature, rank value was restricted to a maximum of 10 i.e. only the 10 potentially best MS/MS spectra per feature were considered for

database search. Spectra were exported as peak lists and database searches were performed using Mascot, OMSSA, and X! Tandem. Next, the search results from different algorithms were combined and filtered at a FDR of 1% using PeptideShaker. Thus generated quality-controlled identification results were then directly re-imported into Progenesis to match quantified features to peptide and protein identifications. Then for each protein, the average of the NAVs (obtained from Progenesis) from the triplicate analyses was calculated to determine the ratios between the scrambled controls and SIL-1 depleted samples. Only proteins that were (i) commonly quantified in all the replicates, (ii) with at least two unique peptides, (iii) an ANOVA p-value of ≤ 0.05 (from Progenesis) and (iv) an average ratio (ΔSIL1/Scr) ≤ 0.667 or ≥ 1.6 (corresponding to 1.6-fold regulation; log2 ratios of +/- 0.65) were considered as regulated.

3.2.15.2.2 Normalized spectral abundance factor (NSAF)

Human RCMH data

In total 16 MS raw files were converted into peak lists (mgf files) using ProteoWizard and were searched separately using Mascot, MS-GF+, and X! Tandem. The identification results from the different search algorithms were imported into PeptideShaker for data interpretation and validation as described above. Next, the 1% FDR quality controlled data was exported and only proteins that were identified with ≥ 1 validated peptide were considered for calculating NSAF value [111a] for each protein. Additionally, the abundance of each protein was normalized with the abundance of glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) - well-known housekeeping protein.

$$(\mathrm{NSAF})\mathrm{k} = \frac{(\mathrm{SpC}/\mathrm{L})\mathrm{k}}{\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{SpC}/\mathrm{L})\mathrm{i}}$$

NSAF = Normalized spectral abundance factor; k = protein; SpC = total number of MS/MS spectra identifying a protein (k); L = length of the protein (k); N = total number of proteins in an experiment.

Human skeletal muscles

A comparative proteomic analysis of human skeletal muscles was not feasible with iTRAQ or precursor area-based label-free quantification approaches due to (i) limited number of samples,

(ii) very low amount of sample material (~5 µg) and (ii) high biological variation between the samples i.e. one clinically very ill patient against two healthy controls. For this reason, NSAF values of the raw MS data were calculated (as described above) in each dataset and the resulting estimated protein abundances were compared between the index patient and two controls datasets to achieve relative quantification. As the data comparison was based on a single shot LC-MS analysis, stringent filtering criteria were applied to increase both quality and confidence of the quantification results. Firstly, only those proteins that were identified in all the three datasets with ≥ 2 validated peptides and spectra (1% FDR) were considered for calculating the NSAF values. Secondly, RSD values were calculated for each protein by taking the NSAF values of the two controls. Thirdly, a ratio (Index/Control) was calculated for each protein by taking the NSAF value of the index patient and an average NSAF value of the controls (see below).

$$\frac{\text{NSAF Index}}{(\text{NSAF Control1} + \text{NSAF Control2})/2}$$

Next, for each protein, the ratio (Index/Control) was normalized by a correction factor obtained by the alignment of TICs of the three samples to compensate for the systematic errors. Lastly, the normalized ratios were log2-transformed and only proteins with a ratio of $\leq$ 0.49 or $\geq$ 2.03 (corresponding to 2-fold regulation; log2 ratios of +/- 1.0) and RSD values ≤ 20% (only controls) were considered as regulated.

3.2.15.3 Targeted - MS data

PRM raw data analysis of SIL1-depleted HEK293 cells was performed using the Skyline software. For both sample sets, triplicate measurements of ΔSIL1_1 and ΔSIL1_2, respectively were compared to the corresponding control (Scr) triplicates as previously described (see Section 3.2.15.2.1). The scheduled targeted MS/MS data of 9 samples was imported into Skyline for visualization and validation. For each peptide/precursor ion (total 74 corresponding to 26 UPR-related proteins), manual validation of MS/MS signals was performed since the peaks generated from the co-eluting peptides with identical m/z values as that of targeted peptide ion could lead to false peak annotations. Hence, an in house generated Q Exactive spectral library which contained the information (a blue print) of the experimental MS/MS spectra of the peptides

that were selected for this assay was used for peak selection and assignment. Furthermore, for a given targeted peptide the corresponding fragment ion peaks which were ambiguous and below the limit of quantification of the instrument (i.e. Q Exactive MS) were discarded. Only those fragment ion peak areas were used for integration (i) that were observed within the predicted retention time windows of the corresponding precursor ion, (ii) with product ion intensities corresponding to the spectral library information and (iii) detected within 10 ppm mass accuracy. Next, the corresponding peptide peak areas were exported to generate protein ratios (ΔSIL1/Scr) and Student's T-Test p-values (two-sample assuming unequal variance) were determined. Only proteins that at least had a reproducible ≥ 1.3-fold (up/down) regulation compared to control and had a Student's T-Test p-values ≤ 0.05 were considered as regulated. However, three out of 26 proteins selected for the PRM-assay method showed inconsistencies in statistical evaluation and had to be removed. The details of the UPR-related proteins used for the PRM-based targeted assay are given in the Appendix 10.8.

4 Results

4.1 *In vitro* protein carbamylation

In aqueous solutions, urea dissociates upon heating and over time. One of its major degradation products is isocyanate, which in buffer solutions (pH between 7.0 - 9.0) containing proteins; covalently reacts mainly with N-termini and epsilon amino group of Lys residues [139] (Fig. 4.1).

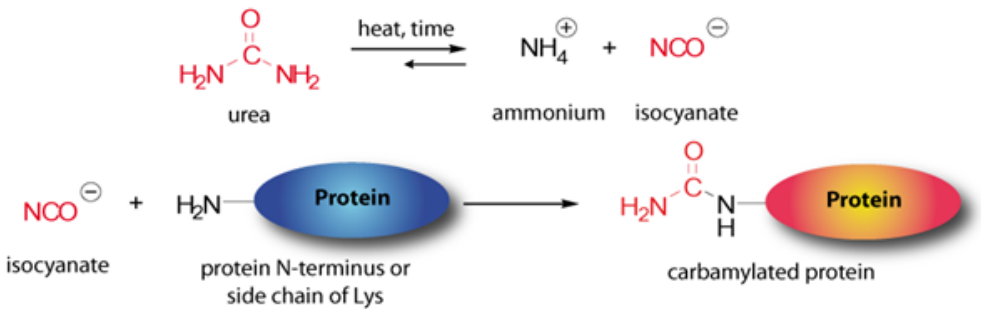


Figure 4.1: Urea-induced carbamylation reaction mechanism in presence of heat (> 37°C) or prolonged exposure with urea in aqueous buffers containing protein/peptides.

In urea-based sample preparation protocols e.g. FASP [120], a systematic study was conducted to evaluate the extent of carbamylation under conditions that are typically used during sample preparation and proteolytic digestion [122] (Table 4.1). For this purpose, a peptide mixture of five model proteins namely, serum albumin, β-casein, β-lactoglobulin, hemoglobin (all bovine) and myoglobin (horse) was generated as described in section 3.2.12.1. By using a stock on the peptide instead of the protein level enabled better inter-comparability and consequently quantification of the extent of carbamylation as this allowed to reduce the technical variability to level of carbamylation and exclude proteolytic digestion as a source of differences between samples. Moreover, this allowed for a more comprehensive study of N-terminal carbamylation compared to the presence of only five accessible N-termini on the protein level [122].

As expected, when treated with harsh conditions (2), nearly all identified peptides were carbamylated (99% N-termini, 90% Lys) (Fig. 4.2 A). Under conditions often used for reduction (3), nearly 13% of peptides were carbamylated at their N-terminus whereas only 4% of Lys side chains were modified. This might be attributed to higher reactivity of primary amines towards isocyanate compared to the epsilon amino group on Lys at basic pH [140]. In many protocols, after

lysis (and sometimes initial Lys-C digestion) urea concentration is lowered to 2.0 M to keep polypeptides in an unfolded state and solubilized during digest. Under these conditions, nearly the same number of carbamylated peptides were identified for both buffers tested i.e. (4) NH_4HCO_3 (21% N-termini, 2.2% Lys) and (5) Tris-HCl (23% N-termini, 1.2% Lys), suggesting that one fifth of N-termini and ~2% of Lys residues can be carbamylated during overnight incubation at 37°C in presence of 2.0 M urea. Upon lowering the urea concentration to 0.1 M in the digestion buffer (6), 2% of N-termini and 0% of Lys residues were modified (Fig. 4.2 A).

Table 4.1: Summary of the different conditions used to assess the degree of carbamylation during common sample processing procedures (3-9) when compared to controls (1-2).

Condition	Buffers	Incubation conditions	Protease	Purpose	Adapted from
1	No treatment (0.1% TFA (v/v))	-	Trypsin	Control	-
2	8.0 M urea, 100 mM Tris-HCl (pH 8.5)	61°C, 15 h	Trypsin	Harsh	-
3	8.0 M urea, 100 mM Tris-HCl (pH 8.5)	56°C, 30 min	Trypsin	Reduction	[127]
4	2.0 M urea, 50 mM NH_4HCO_3 (pH 7.8)	37°C, 15 h	Trypsin	Digestion	[128]
5	2.0 M urea, 100 mM Tris-HCl (pH 8.5)	37°C, 15 h	Trypsin	Digestion	[129]
6	0.1 M urea, 50 mM NH_4HCO_3 (pH 7.8)	37°C, 15 h	Trypsin	Digestion	-
7	0.2 M GuHCl, 50 mM NH_4HCO_3 (pH 7.8)	37°C, 15 h	Trypsin	Digestion	-
8	8.0 M urea, 100 mM Tris-HCl (pH 8.5)	20°C, 18 h	Trypsin	Digestion	[120]
9	8.0 M urea, 100 mM Tris-HCl (pH 7.9) and 2.0 M urea, 100 mM Tris-HCl (pH 7.9)	20°C, 18 h and 20°C, 4 h	Lys-C Trypsin	Digestion Digestion	[120]

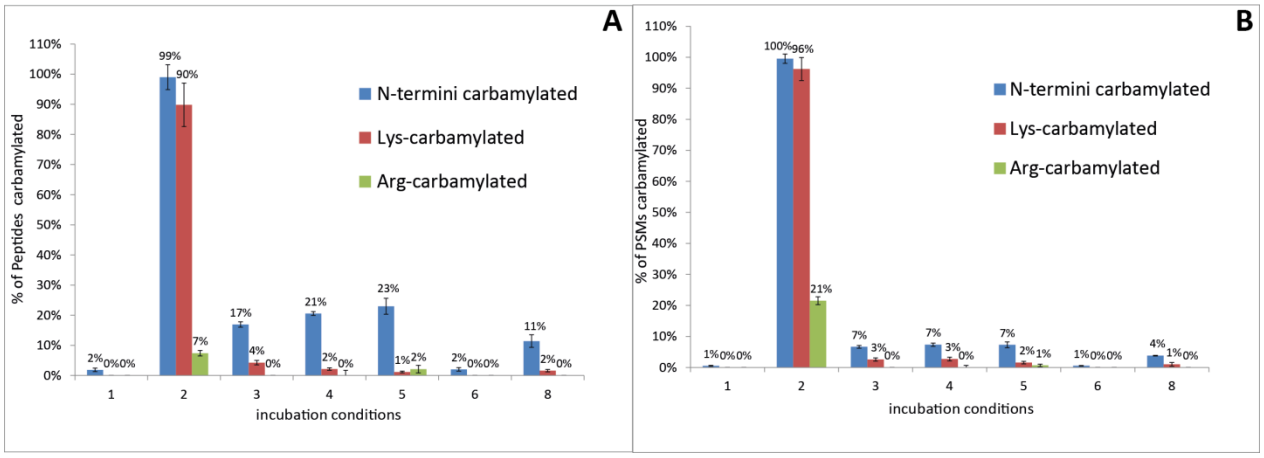


Figure 4.2: Share of carbamylated peptides (**A**) and PSMs (**B**) identified after treating the peptide mixture (5 proteins) with different incubation conditions as summarized in Table 4.1 [122].

In addition, a buffer with GuHCl was evaluated as an alternative chaotropic agent; as expected carbamylation could not be detected due to the absence of urea, albeit, primary amines are potentially prone to guanidinylation in presence of GuHCl. However, overnight incubation at 37°C with GuHCl (7) did not result in detectable guanidinylation of N-termini and Lys side chains.

Next, (8) the conditions used for Lys-C digestion were evaluated by incubating the peptide mixture in 8.0 M urea for 18 h at 20°C, thus detecting 11% of N-termini and 2% of Lys residues carbamylated (Fig. 4.2 A). Notably, when prepared freshly, no differences could be detected between the 4-month and the 8-year old urea, suggesting that urea degradation occurs only in aqueous solutions.

Because of the relatively high share of urea-induced carbamylation with Lys-C, the conditions often used during two-step digestion procedure (9) involving initial proteolysis in 8.0 M urea (similar to 8), followed by 4-fold dilution and digestion with trypsin at 20°C for 4 h were evaluated. For this, human fibroblasts were lysed with 0.1% Rapigest and a two-step digestion was performed with Lys-C and trypsin (see Section 3.2.12.2), once in the presence of urea (urea-based) and once in Rapigest (urea-free). After the database searches of LC-MS data, the number of identified non-carbamylated PSMs was nearly identical in both cases i.e. 12,982 in urea-based and 12,321 in urea-free samples (Table 4.2).

Table 4.2: Summary of two-step digest condition (9) performed with Lys-C and trypsin using fibroblast lysates.

Condition	number of proteins identified with ≥ 1 unique peptide	number of peptides identified	number of PSMs	% of carbamylated peptides (N-termini, Lys and Arg)
Urea-based	1578 ± 17	6465 ± 330	12982 ± 720	0.3% ± 0.1%
Urea-free (Rapigest)	1547 ± 13	5806 ± 252	12321 ± 460	0.2% ± 0.1%

Nevertheless, as an example for the potential severity of the problem, the occurrence of urea-induced carbamylation was evaluated in a recently published study that aimed to investigate the N-terminal proteome of *Escherichia coli* (prepared using urea) in a single 6 h LC-MS run [128]. For this, the MS raw data was re-searched using the same search parameters that were used for the carbamylation experiment in this work. Remarkably, the database searches revealed that at 1% FDR, 27.4% of the identified peptides (2,597 out of 9,481) and 25.7% of the PSMs (4,103 out of 15,948) were indeed carbamylated preferentially on the N-termini (~24% compared to ~7% of Lys residues) [122].

4.2 Investigation of clinically unaffected cell types - Human

4.2.1 MSS fibroblasts

The selective vulnerability of SIL1-deficiency in certain tissues and organs of MSS patients has been confirmed by different research groups [5, 18, 41]. However, ultra-structural investigations (i.e. TEM, Fig. 4.3) on MSS-patient derived fibroblasts revealed morphological alterations [141] suggesting that loss of functional SIL1 also has an effect on tissues or cellular populations that are apparently not MSS-vulnerable. One of these findings include dilated ER lumen suggesting an increased burden on the ER due to the accumulation of most likely, unfolded or misfolded proteins. Furthermore, the presence of vacuoles filled with electron-dense membranous coils (so called myelin-like material) adjacent to the ER indicates proteolytic degradation (Fig. 4.3). One of the major tasks of this work was to further investigate these findings and to see whether one can identify the molecular reasons for these changes on the proteome level and deduce what could be a partially compensating effect that renders these tissues unaffected in MSS patients. For this, iTRAQ-based (4-plex) proteomics analysis was performed with the cultured fibroblasts derived from two MSS-patients (family MSS2) and their respective healthy controls. The LC-MS analysis of the 24-OFFGEL fractions led to the quantification of 2,993 proteins (≥ 2 unique peptides, 1% FDR). Out of these, 136 proteins showed altered levels (set criteria: an average MSS/Healthy ratio of ≤ 0.5 or ≥ 2.0 for down-/upregulation and RSD ≤ 20%) and among which 57 proteins were down- and 79 proteins were upregulated when compared to healthy fibroblasts. Details of altered proteins are given in the Appendix 10.3.

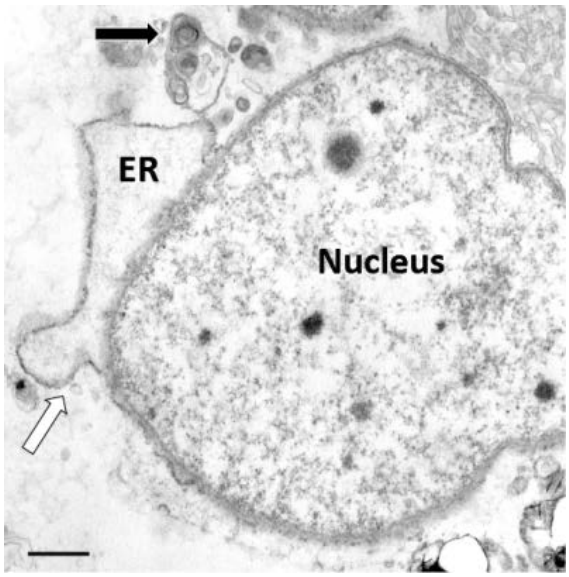


Figure 4.3: TEM findings in MSS2 fibroblasts showing dilated ER-lumen and ER membrane with a high density of ribosomes (white arrow) and abnormal membranous coils (black arrow). Scale 1 µm.

SIL1 - previously described as a low abundant ER-resident co-chaperone [142], was one of the downregulated proteins (Fig. 4.4). Moreover, this finding is in line with the mutation analysis of the MSS2 case. Genetic studies of MSS2 patient identified missense or nonsense type mutation in *SIL1* leading to an expression of SIL1 transcripts (mRNA) lacking the complete exon 6 (so-called exon skipping). Despite the absence of exon 6, the SIL1 protein is still translated, but with an in-frame deletion of 64 amino acids (position Ala152 to Gln215) and it is predicted to be present in low levels in MSS2 patients [40] (Fig. 4.5). Furthermore, one of the major interaction sites of SIL1 with BiP is composed of these 64 amino acids [41b]. As a result, even if present in low amounts, the truncated SIL1 protein will not be able to bind the ADP-BiP-substrate complex and catalyze the release of ADP.

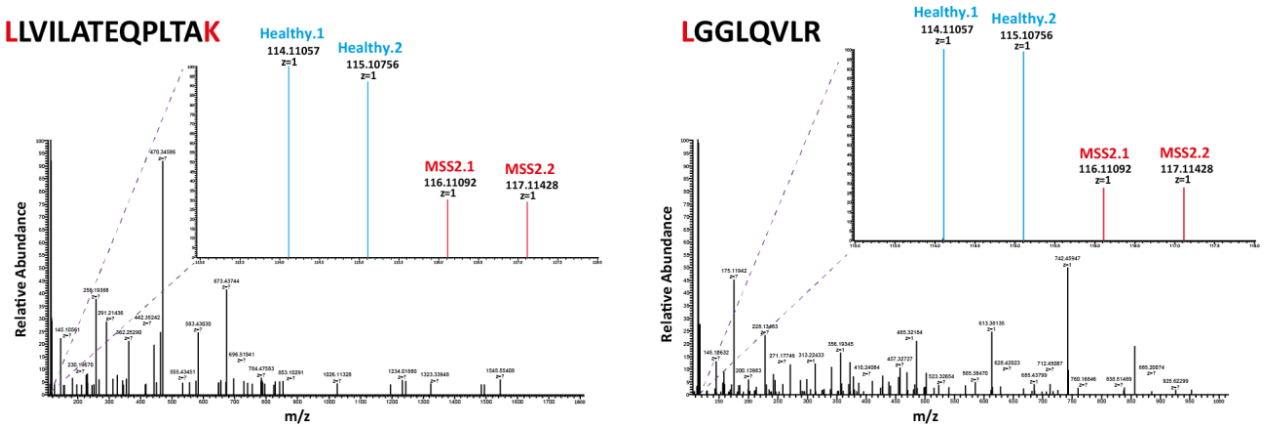


Figure 4.4: Product ion (MS/MS) spectra acquired in the Orbitrap mass analyzer (R = 7,500 FWHM at m/z 400) of two iTRAQ 4-plex labeled peptides that belong to SIL1. The amino acids marked in red indicate the sites of attachment of the isobaric tags. As shown, the relative intensities of the reporter ions of the two MSS2 patients' replicates reflect the relatively reduced abundance of SIL1 with respect to healthy controls. Notably, the real downregulation of SIL1 might be higher than the here indicated ~3-fold, as reporter ion based quantification generally tends to underestimate regulation levels.

>sp|Q9H173|SIL1_HUMAN Nucleotide exchange factor SIL1 OS=Homo sapiens GN=SIL1 PE=1 SV=1
MAPQSLPSSRMAPLGMLLGLLMAACFTFCLSHQNLKEFALTNPEKSSTKETERKETKAEE
ELDAEVLEVFHPTHEWQALQPGQAVPAGSHVRLNLQTGEREAKLQYEDKFRNNLKGKRLD
INTNTYTSQDLKSALAKFKEGAEMESSKEDK**ARQAEVKRLFRPIEELKKDFDELNVVIET**
DMQIMVRLINKFNSSSSSLEEKIAALFDLEYYVHQMDNAQDLLSFGGLQVVINGLNSTEP
LVKEYAAFVLGAAFSSNPKVQVEAIEGGALQK**LLVILATEQPLTAK**KKVLFALCSLLRHF
PYAQRQFLK**LGGLQVLR**TLVQEKGTEVLAVRVVTLLYDLVTEKMFAEEEAELTQEMSPEK
LQQYRQVHLLPGLWEQGWCEITAHLLALPEHDAREKVLQTLGVLLTTCRDRYRQDPQLGR
TLASLQAEYQVLASLELQDGEDEGYFQELLGSVNSLLKELR

Figure 4.5: FASTA sequence of human SIL1 protein (http://www.uniprot.org/uniprot/Q9H173.fasta). The motif i.e. Ala152 to Gln215 (amino acids indicated in red), which is responsible for the interaction between SIL1 and BiP is absent in the MSS2 family patient due to the lack of exon 6 in SIL1 mRNA transcripts. The amino acid sequences highlighted in green were detected and quantified by iTRAQ-based LC-MS analysis (see Figure 4.4).

The MSS2/Healthy ratios of BiP and GRP170 showed a marginal upregulation (both ~1.6) and thus these two candidates did not pass the strict criteria for differentially regulated proteins.

Notably, nine out of 136 altered proteins have been associated with various autosomal recessive disorders and MSS being one of them according to the information listed in UniProtKB (http://www.uniprot.org) as of December 2015. The following protein-related data which are not derived from UniProtKB are marked by their references. Among the upregulated proteins, alpha-crystallin B chain (*CRYAB*) has been related to cataracts and muscular dystrophy and mitochondrial superoxide dismutase [Mn] (*SOD2*) was reported to cause retinopathy (an eye disease). Furthermore, bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 1 (*NDST1*) is involved in mental retardation, whereas bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 2 (*PAPSS2*) and basement membrane-specific heparan sulfate proteoglycan core protein (*HSPG2*) have been associated with skeletal muscle abnormalities. Interestingly, apart from retinopathy, the other phenotypes are similar to the prominent clinical findings in MSS patients [41a]. Notably, heme oxygenase (*HMOX1*) and heat shock protein beta-6 (*HSPB6*) - both involved in the oxidative stress response pathway, showed increased abundances suggesting the ER-stress and subsequent activation of the UPR pathway [143]. Furthermore, ubiquitin carboxyl-terminal hydrolase isozyme L1 (*UCHL1*), which is involved in processing of the ubiquitin precursors as well as ubiquitinated proteins showed increased levels in MSS2 fibroblasts, indicating a potential triggering of the ERAD leading to proteasome-mediated ubiquitin-dependent protein catabolic process. Remarkably, *UCHL1* has been related to two neurodegenerative disorders, one of them being the Parkinson's disease 5 [144] and the other one is childhood-onset neurodegeneration with optic atrophy (NDGOA).

The imbalance between the rates of protein production and protein degradation could be one of the reasons for aggregate formation. Carboxypeptidase M (*CPM*) - a protein processing enzyme that is involved in the membrane-localized degradation of extracellular proteins was downregulated in the MSS2 fibroblasts. Similarly collectin-12 (*COLEC12*), which is presumed to play a role in clearance of amyloid β peptides in the Alzheimer's disease also showed reduced abundance in MSS2. These results are in accordance with the TEM findings that detected dense membranous material in the vicinity of the ER and the nucleus (Fig. 4.3).

4.2.2 MSS lymphoblastoid cells

Application of iTRAQ 8-plex technology, high-pH RP fractionation and LC-MS analysis led to the quantification of 4,858 proteins (≥ 2 unique peptides, 1% FDR) in the LCs obtained from four individual patients i.e. MSS24, MSS32, MSS33 and MSS94 along with four different healthy controls. The chemical labeling strategy revealed 112 proteins with significantly altered abundances (set criteria: Student's T-Test p-value ≤ 0.05, an average MSS/Healthy ratio of ≤ 0.75 or ≥ 1.32 for down-/upregulation) in MSS-LCs of which 48 proteins were down- and 64 proteins were upregulated. Details of altered proteins are given in the Appendix 10.4.

As already seen in MSS-derived fibroblasts, SIL1 was again among the downregulated proteins in MSS-derived LCs. Thereby, the genetic cause in all MSS patients (MSS24, MSS32, MSS33 and MSS94 [41a]) is in accordance with the detection of residual - most likely non-functional SIL1 protein level in this sample set (Fig. 4.6 A) and furthermore, these results are in line with the WB analysis (Fig. 4.6 B).

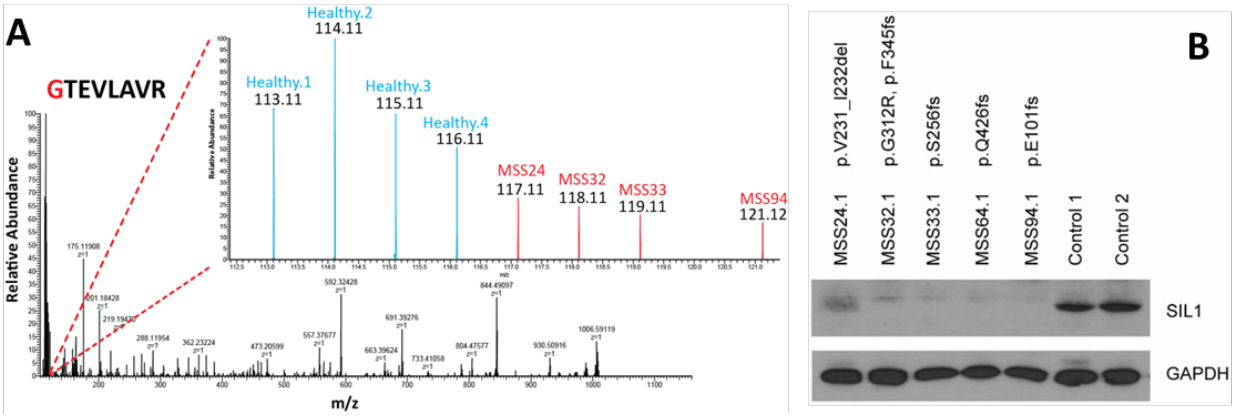


Figure 4.6: (A) Product ion (MS/MS) spectra acquired in the Orbitrap (R = 7,500 FWHM at m/z 400) of an iTRAQ 8-plex labeled SIL1 peptide. The amino acid marked in red indicates the site of attachment of the isobaric tag. (**B**) Western Blot analysis performed with five different MSS patients' and two healthy controls using an anti-SIL1 antibody (top) and an anti-GAPDH antibody (as loading control, bottom). The image is adapted from [41a] which shows that SIL1 mutations result in substantially decreased SIL1 levels depending on the type of mutation in patients MSS24, MSS32, MSS33, MSS64 and MSS94, respectively compared to the levels in controls. "The seemingly "milder" mutations [small in-frame deletion (MSS24.1), missense mutation (MSS32.1) and frameshift mutation in the last exon (MSS64.1)] have similar effects as truncating mutations that are expected to lead to nonsense mediated messenger RNA decay (MSS33.1 and MSS94.1)" [41a].

However, the expression levels of both BiP and GRP170 were almost stable (MSS/Healthy ratio for both = 0.8). Nevertheless, similar to the MSS-fibroblasts, alterations in some of the subcellular compartments were detected in this clinically unaffected cellular population as well. It is known that mitochondrial dysfunction caused by accumulation of misfolded proteins is

linked to various neurodegenerative and neuromuscular disorders in man and mouse [35, 145]. Proteomic profiling of MSS-LCs revealed increased abundances of mitochondrial proteins which promote apoptosis i.e. cytochrome C (*CYCS*) and apoptosis-inducing factor 2 (*AIFM2*). Besides, reduced abundances of mitochondria-associated carnitine O-palmitoyltransferase 1, liver isoform (*CPT1A*), peptide chain release factor 1-like (*MTRF1L*), TOM1-like protein 2 (*TOM1L2*) and thioredoxin reductase 2 (*TXNRD2*) indicate abnormal functioning of this subcellular organelle. Nevertheless, upregulation of mitochondrial short/branched chain specific acyl-CoA dehydrogenase (*ACADSB*), phospholipid hydroperoxide glutathione peroxidase (*GPX4*) and peroxiredoxin-5 (*PRDX5*) is in agreement with the activation of pro-survival/mitoprotective strategies in SIL1-deficient LCs to avoid cell death. In addition, downregulation of inhibitor of nuclear factor kappa-B kinase-interacting protein (*IKBIP*), which is a target of cellular tumor antigen (*TP53*) with pro-apoptotic function indicate negative regulation of apoptosis.

Although the ER can form intracellular networks independently of cytoskeletal structures, integrity and distribution of this compartment in mammalian cells are influenced by cytoskeletal components [146]. Remodeling of the cytoskeleton affecting both actin filaments and microtubules upon loss of functional SIL1 in LCs is shown by downregulation of myristoylated alanine-rich C-kinase substrate (*MARCKS*) and upregulation of tyrosine-protein kinase (*HCK*), fermitin family homolog 2 (*FERMT2*), kinesin-like protein (*KIF21A*), profilin-2 (PFN2), protein XRP2 (*RP2*) and septin-1 (*SEPT1*).

Identical to the MSS-derived fibroblasts, proteins that are involved in the ubiquitin-proteasome pathway e.g. ankyrin repeat domain-containing protein 13A (*ANKRD13A*) and ubiquitin-conjugating enzyme E2 variant 3 (*UEVLD*) were upregulated in the MSS-LCs indicating activation of the ERAD pathway. Furthermore, cathepsin B (*CTSB*) - a thiol protease, which participates in intracellular degradation and turnover of proteins showed increased abundance suggesting the role of *CTSB* in clearing misfolded proteins. This is in line with the TEM findings of electron-dense autophagy material observed in the MSS-LCs (Fig. 4.7 C, D).

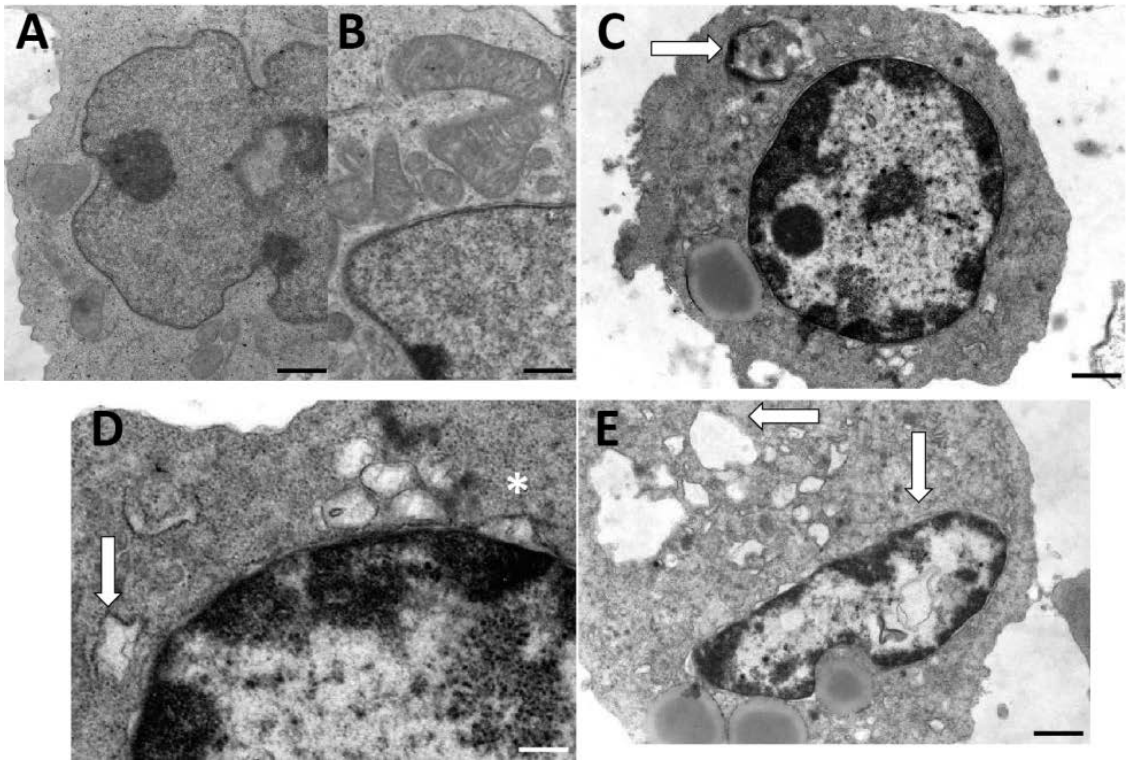


Figure 4.7: TEM findings in control LCs (**A, B**) and LCs derived from MSS patients (**C** - **E**). (**A, B**) Regular organelle structures in LC derived from healthy controls. Scale bars in **A** = 5 µm; in **B** = 2 µm. (**C**) Intense coloration of the nucleus and accumulation of electron-dense material (white arrow) in a LC derived from a MSS patient. Scale bar = 7.5 µm. (**D**) Detailed magnification of (**C**) emphasizes widened rough ER (white arrow) and outfoldings of the lifted-off nuclear envelope (asterisk) as well as accumulation of vesicular structures partially filled with electron-dense material. Scale bar = 1 µm. (**E**) TEM data pf another MSS-patient derived LC presenting with in nuclear and cytosolic alterations (white arrows). Scale bar = 7.5 µm. The TEM studies were performed at the Institute of Neuropathology, Aachen.

SIL1 is ubiquitously expressed in the brain and its loss leads to cerebellar ataxia - a clinical hallmark of MSS. Notably, ataxin-10 (*ATXN10*) is necessary for the survival of cerebellar neurons and induces neuritogenesis by activating the Ras-MAP kinase pathway. Upregulation of *ATXN10* in MSS-LCs might indicate a compensatory strategy to antagonize Purkinje cell degeneration in SIL1-deficient cerebella [147]. Moreover, *ATXN10* plays a role in the maintenance of critical intracellular glycosylation levels and homeostasis. Similarly, galectin-1 (*LGALS1*), which belongs to the family of β-galactoside-binding proteins showed increased abundances in SIL1-deficient LCs. *LGALS1* has been reported to be involved in various cell defense mechanisms including pro-survival and establishment of homeostasis [148]. In addition, it was shown that galectin-1 has an indirect neuroprotective function in immune mediated inflammatory neurodegenerative disorders e.g. multiple sclerosis [149].

4.3 Investigation of clinically affected tissues - Mouse

To gain insights into this pathological condition of SIL1-loss on the global level, proteomic profiling of Sil1-deficient affected tissues i.e. cerebellum and skeletal muscles derived from the *woozy* mice and the respective wild type littermates was performed.

4.3.1 *Woozy* cerebellum

The cerebellum also known as the "little/small brain" is located at the back of the mammalian brain and it constitutes ≥ 50% [150] of the total number of neurons - the basic working units of the brain. Its main functions are maintaining balance and posture, coordinating motor activities and learning new cognitive skills. Cerebellar damage is usually manifested in the form of ataxia i.e. loss of motor coordination, which is a prominent overlapping symptom between *SIL1*/*Sil1*-mutant man and *woozy* mouse.

Similar to MSS-LCs, iTRAQ 8-plex labeling in combination with RP pH 6.0 fractionation and LC-MS/MS analysis were performed using three biological replicates of *woozy* and wild type (26-weeks old, female) cerebella. This led to the quantification of 3,580 proteins (≥ 2 unique peptides, 1% FDR) wherein, ~3% (112) were found to be significantly altered (set criteria: Student's T-Test p-value ≤ 0.05, an average *woozy*/wild type ratio of ≤ 0.74 or ≥ 1.34 for down-/upregulation). Among these 112 proteins, ~79% (88) were down- and ~21% (24) were upregulated. Details of altered proteins are given in the Appendix 10.5.

Remarkably, BiP and its alternate NEF i.e. Grp170, showed relatively stable protein abundances (i.e. *woozy*/wild type ratios of 0.95 and 0.98, respectively) in the *woozy* with respect to wild type littermates, whereas Sil1 could not be quantified. This result however, is in contrast to the findings of Zhao et al., [47] who reported increased levels of BiP in the Purkinje cells of *Sil1*-mutant animals. The plausible reason for the discrepancy between the two datasets can be due to the fact that Zhao and colleagues performed their studies only on the Purkinje cell population. Whereas, the complete cerebellum of *woozy* was used for proteomics analysis, which comprises different cell types. Furthermore, one can speculate the presence of pro-survival mechanisms in other cell populations of the cerebellum that might compensate for Sil1-loss in the *woozy*. To support this presumption, upregulation of certain proteins that are known for their cell survival

(or cytoprotective) capabilities were found in the proteomics data of *woozy* mice. For instance, mast/stem cell growth factor receptor Kit (*Kit*) protein, which plays an essential role in the regulation of cell survival and proliferation, showed upregulation in *woozy* mice. It is well-known that oxidative stress is among the main factors in causing cellular damage. In this regard, upregulation of cytosolic proteins, such as nicotinate phosphoribosyl transferase (*Naprt*) and peroxiredoxin-6 (*Pdrxn6*), which play a protective role against oxidative injury, indicate an activation of defense mechanisms in Sil1-deficient cerebellar neuronal cells. Notably, increased abundance of caspase-3 (*Casp3*), which is involved in the execution of glial cell apoptosis, was found in the neuronal cells of the *woozy*. This finding consolidates the TEM findings (Fig. 4.8 A) that detected the presence of apoptotic material in the nerve cells and thus complementing morphological and proteomics analyses. In contrast, inositol 1,4,5-trisphosphate receptor type 1 (*Itpr1*) - an ER-membrane bound intracellular ligand-gated calcium channel that plays a role in the ER-stress induced cell death showed decreased levels in the *woozy* indicating a negative regulation of apoptosis.

Whereas increased levels of glial fibrillary acidic protein (*Gfap*) were found in the *woozy*. *GFAP* is one of the most widely used markers of neurologic damage in humans [151]. Interestingly, 20 proteins that are associated with intra-/extracellular calcium (Ca^{2+}) regulation, Ca^{2+} storage and signaling were found to be downregulated in *woozy*. Despite the lack of evidence that supports the direct interaction between SIL1 and Ca^{2+}, the altered levels of the latter can be attributed to the dysfunctional SIL1-BiP machinery leading to disturbed cellular homeostasis [34c]. These include (i) Ca^{2+} binding proteins: calcium-dependent secretion activator 2 (*Cadps2*) and calbindin (*Calb1*), (ii) Ca^{2+} dependent ion channels: calcium/calmodulin-dependent protein kinase type IV (*Camk4*) and voltage-dependent calcium channel subunit alpha-2/delta-2 (*Cacna2d2*), (iii) proteins involved in Ca^{2+} homeostasis: regulator of microtubule dynamics protein 3 (*Rmdn3*). Remarkably, *Calb1* - a Purkinje cell marker, was also downregulated in the Zhao et al., dataset.

Cell-cell communication is mainly carried out by signaling pathways involving receptors and ion channels present in the plasma membrane [152]. Notably, protein kinase C gamma type (*Prkcg*) is solely expressed in the brain and the spinal cord and it is the most abundant isozyme of the protein kinase C family in the cerebellum. *Prkcg* is activated by fatty acids and plays diverse roles in cellular signaling events of the neurons in a Ca^{2+}-dependent manner [153] including

regulation of the neuronal receptors e.g. glutamate receptor 4 (*Gria4*) and cell survival after cerebellar ischemia [154]. Owing to its important role in the regulation of neuronal signaling pathways, *Prkcg*-deficient mice manifest symptoms of mild deficits in spatial and contextual learning and impaired motor coordination due to disturbances in the development process of the Purkinje cells [153]. The low abundance level of *Prkcg* therefore correlates with the observed neurodegeneration in *woozy*. Moreover, Ca^{2+} also plays an important role in the transmission of neurotransmitters at the neuromuscular junction - a chemical synapse formed by the contact between a motor neuron and a muscle fiber, by regulating the voltage-dependent ion channels [155]. The relatively low abundances of voltage-dependent P/Q-type calcium channel subunit alpha-1A (*Cacna1a*) and sarco-/endoplasmic reticulum calcium ATPase 2 (*Atp2a2*) might indicate neuromuscular degeneration in the SIL1-mutant *woozy* mouse and MSS patients.

The cytoskeleton of a nerve cell is typically elaborated due to the axon and the dendrite - parts of a neuron that specialize in exchange of information (or impulses) between neurons at the synapse. Notably, cytoskeletal abnormalities caused due to the altered intracellular Ca^{2+} levels have been associated with aging and certain neurodegenerative disorders [156]. Besides *Camk4*, calcium/calmodulin-dependent protein kinase type II subunit beta (*Camk2b*), which plays an important role in the cytoskeletal reorganization of neuronal cells, was also downregulated in *woozy* suggesting a negative effect on the structural integrity of these cells. Additionally, SH3 and multiple ankyrin repeat domains protein 1 (*Shank1*) and 2 (*Shank2*) that are involved in the structural and functional organization of the dendrites particularly at the synaptic junction, showed reduced abundances in *woozy*. Further, delphilin (*Grid2ip*), a postsynaptic scaffolding protein present at the Purkinje cell synapse was also downregulated, which clearly shows the degeneration of this neuronal cell population as a consequence of Sil1-loss.

In addition, the TEM findings in the neocerebellum (Fig. 4.8 A-D) of *woozy* mouse (26-weeks old) revealed electron-dense material present in the neuronal cells indicative of protein aggregates. The presence of lysosomal autophagic material (Fig. 4.8 A) in the nuclear and perinuclear regions indicate activation of the ER-stress induced autophagy-lysosomal protein degradation. This finding is in accordance with the results from Zhao et al., suggesting that loss of the Purkinje cells causes early onset of ataxia in *woozy* animals. In addition to the above findings,

damaged mitochondria (which usually trigger apoptosis) were also detected in the nerve cells (4.8 D).

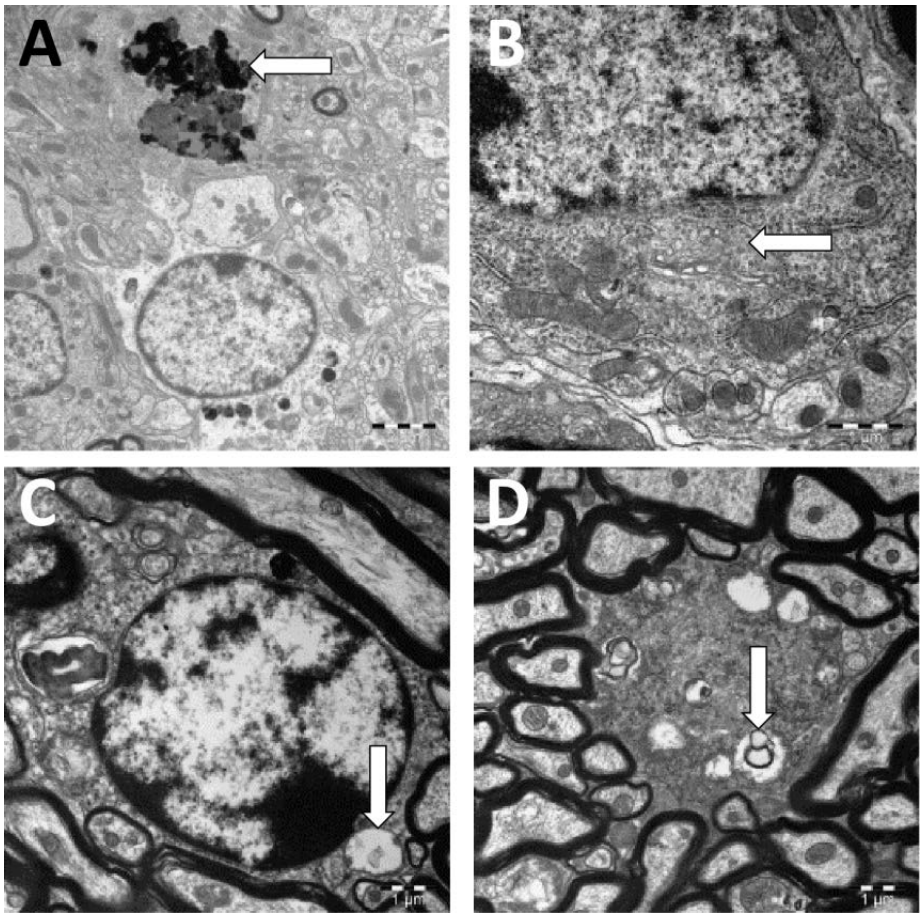


Figure 4.8: TEM findings in neocerebellum of *woozy* mouse at the age of 26 weeks: (**A**) Accumulation of electron-dense material and lysosomes in the vicinity of a nucleus (white arrows). (**B**) Proliferation of vesicular structures partially filled with electron-dense material within a glial cell (white arrows). (**C**) Degeneration of small myelinated axons (white arrows) in the proximity of a neuronal nucleus. (**D**) Mitochondrial degeneration (white arrow) in an apoptotic neuronal cell surrounded by small and large myelinated axons. Scale bar = 1 µm. The TEM studies were performed at the Institute of Neuropathology, Aachen.

4.3.2 *Woozy* skeletal muscles

The mammalian skeletal muscle consists of elongated, multinucleated, transversely striated fibers that are typically attached to bones and tendons. Its main function is contraction and relaxation, which provides stability and movement of the body. By employing an identical iTRAQ-based LC-MS methodology as above, a total 2,055 proteins (≥ 2 unique peptides, 1% FDR) were quantified of which 64 proteins were significantly differentially regulated (set criteria:

Student's T-Test p-value ≤ 0.05, an average *woozy*/wild type ratio of ≤ 0.70 or ≥ 1.43 for down-/upregulation) in *woozy* mice. Details of altered proteins are given in the Appendix 10.6.

Notably, ~83% (53) of the altered proteins were up- and the remaining 11 proteins were downregulated. This relatively lower number of the total quantified proteins compared to the primary cell lines data (i.e. MSS fibroblasts and LCs) can be attributed to (i) the complex nature of the skeletal muscle itself [157] and (ii) the presence of highly abundant cytoskeletal proteins i.e. actin and myosin. Not surprisingly, 46% of all quantified and validated (1% FDR) PSMs belong to these two candidates (including different types and their isoforms) in this iTRAQ dataset. Based on the NSAF value calculation, this corresponds to 34% of estimated relative protein abundance of various forms of actin and myosin. Nevertheless, some of the key UPR-related proteins i.e. BiP (*Hspa5*), calnexin (*Canx*) and protein disulfide isomerase-6 (*Pdia6*) (Fig. 4.9) could still be detected demonstrating the robust combination of iTRAQ technology, high-pH RP fractionation and sensitivity of the LC-MS analysis. These results are in agreement with the WB findings that showed elevated levels of the UPR pathway proteins (Fig. 4.10). Besides, increased levels of proteins that are involved in (i) the ERAD pathway: transitional endoplasmic reticulum ATPase (*Vcp*), dnaJ homolog subfamily B member 6 (*Dnajb6*) and (ii) proteolysis: beclin-1 (*Becn1*) and TAR DNA-binding protein 43 (*Tdp43*) could be detected in the WB analysis (Fig. 4.10). The *woozy*/wild type ratio of Grp170 also showed an upregulation in iTRAQ dataset (~1.4) however, it did not pass the strict criteria for differentially regulated proteins. Notably, Grp170 (*Hyou1*), the ER-resident chaperone/NEF for BiP, indeed showed moderately increased levels in Sil1-mutant mice. In contrast, overexpression of *Hyou1* in skeletal muscles was reported to cause severe myopathic changes in mice [158] and therefore its upregulation might be compromised to such levels as to avoid any negative effects causing further muscular damage [35].

Moreover, upregulation of cytosolic ERAD-assisting chaperones i.e. heat shock protein beta-1 (*Hspb1*) and beta-7 (*Hspb7*) was seen in the iTRAQ data of *woozy*. Interestingly, alpha-crystallin B chain (*Cryab*) (involvement in cataracts and muscular dystrophy), which showed increased abundance in the MSS2-fibroblasts was also among the upregulated candidates in the *woozy* mice.

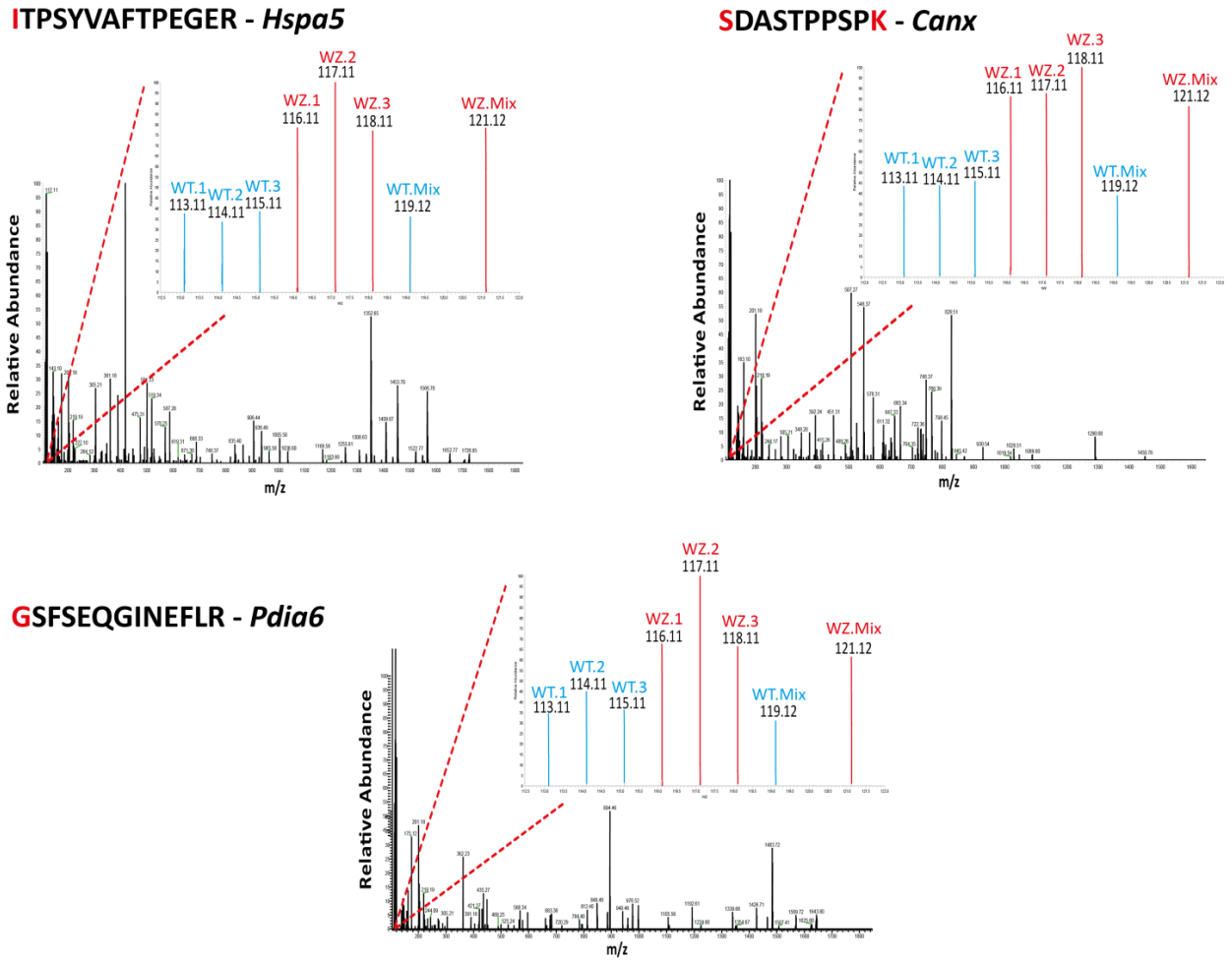


Figure 4.9: Representative product ion (MS/MS) spectra acquired in the Orbitrap (R = 17,500 FWHM at m/z 200) of iTRAQ 8-plex labeled peptides which belong to BiP, calnexin and protein disulfide isomerase A6, respectively. The amino acid(s) marked in red indicates the site of attachment of the isobaric tag. **Abbreviations:** WT = wild type, WZ = *woozy*, WT Mix and WZ Mix = pooled samples of three biological replicates of each type, respectively.

Proteins that are involved in the cytoskeletal organization showed increased abundances implying possible structural myopathic changes in the muscle fibers of *woozy*. These include: coronin-1A (*Coro1a*), tubulin beta-6 chain (*Tubb6*), profilin-1 (*Pfn1*), cofilin-1 (*Cfl1*) and xin actin-binding repeat-containing protein 1 (*Xirp1*). *Coro1a* and *Pfn1* are actin-binding proteins, and *Tubb6* is a major part of microtubules, which are filamentous intracellular structures responsible for various kinds of movements in all eukaryotic cells. Moreover, *Xirp1*, which protects actin fibers from depolymerization, has been associated with the abnormal thickening of cardiac muscle in mice.

It is well-known that Ca^{2+} plays a major role in muscle contraction - being the main signaling and regulatory molecule in the muscle fibers [159]. Calsequestrin-2 (*Casq2*), a Ca^{2+}-binding protein

(which also acts an internal calcium store in the sarcoplasmic reticulum of muscle cells) showed increased abundance in *woozy*. *Casq2* regulates the release of lumenal Ca^{2+} through the calcium release channel ryanodine receptor 2 (*Ryr2*) and plays an important role in triggering muscle contraction as well as in excitation-contraction coupling in cardiac muscles [160]. Similarly, other Ca^{2+} binding protein proteins i.e. protein S100-A13 (*S100a13*), protein S100-A6 (*S100a6*), annexin A1 (*Anxa1*) and annexin A4 (*Anxa4*) showed high levels in *woozy* mice when compared to their wild type littermates. Upregulation of these Ca^{2+} associated proteins clearly indicate activation of the UPR pathway due to altered Sil1-BiP machinery or might reflect a cellular attempt to antagonize muscle fiber degeneration by forced contraction.

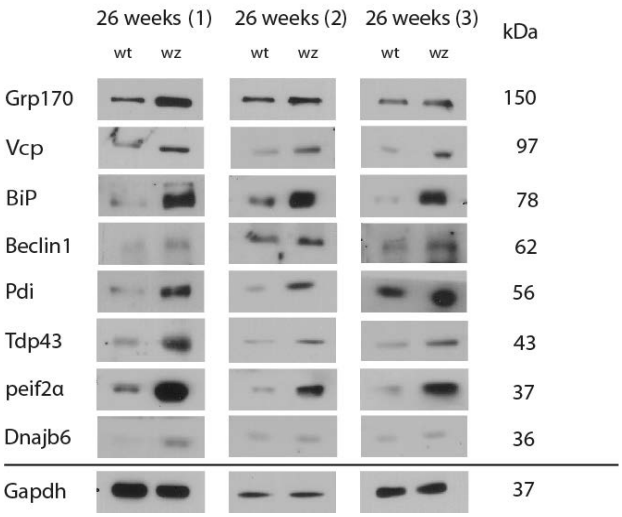


Figure 4.10: Comparative WB analysis of skeletal muscles derived from *woozy* and wild type littermates (female, 3 biological replicates each). Immunoblots indicate elevated levels of proteins that are associated with (i) the UPR pathway (Grp170, BiP, Pdi, peif2α), (ii) the ERAD pathway (Dnajb6, Vcp) and (iii) the proteolysis pathway (Beclin1, Tdp43) in *woozy* mice. Gapdh was used as loading control. The WB analysis was performed at the Institute of Neuropathology, Aachen.

Galectin-3 (*Lgals3*), which belongs to the same family of lectins as galectin-1 (upregulated in the MSS-LCs), plays an essential role in the regulation of cellular homeostasis. This protein was shown (i) to modulate cell growth, (ii) to control the cell cycle, and also (iii) involvement in the regulation of apoptosis [161]. Owing to *Lgals3* role in cell death inhibition, its upregulation in the *woozy* might indicate an activation of pro-survival mechanism in the Sil1-deficient *woozy* muscles. Besides, overexpression of elongation factor 1-alpha 1 (*Eef1a1*), which is involved in protein biosynthesis, has been reported to play a possible protective role under ER-stress caused due to aggregation of un-/misfolded proteins [162]. Moreover, nestin (*Nes*), which is required for brain and eye development and survival [163] was also overexpressed in *woozy*, which is in accordance with the fact that apart from skeletal muscle, the brain and the eyes are severely affected by SIL1-loss in man. Similarly, mesencephalic astrocyte-derived neurotrophic factor (*Manf*), which has been reported to selectively promote the survival of dopaminergic neurons by inhibiting the ER-stress induced cell death, was also upregulated.

4.4 Proteomic profiling of an index patient - Human

Owing to the high sensitivity of MS and currently available quantification strategies, it was evaluated whether a comparative proteomics analysis can be conducted with a low amount of starting material (~5 µg of protein) obtained from a skeletal muscle biopsy of an index patient. After nano-LC-MS analysis, the NSAF values calculation and data normalization (see Section 3.2.15.2.2), a total 95 proteins were relatively quantified (set criteria: an average index/control ratio ≤ 0.49 or ≥ 2.03 for down-/upregulation, RSD ≤ 20%). Notably, ~93% (88) of the overall quantified proteins were up- and 7 proteins were downregulated. Details of altered proteins are given in the Appendix 10.7.

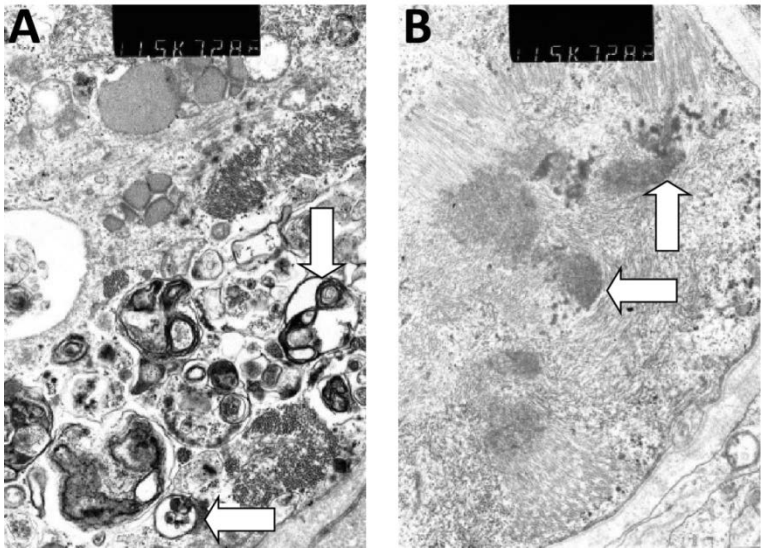


Figure 4.11: TEM findings on the muscle biopsy derived from the index patient. **(A)** Huge sub-sarcolemmal accumulation of vacuoles filled with electron-dense material most likely corresponding to protein aggregates (white arrows) and thus indicating a protein folding disorder as the cause of myopathy. **(B)** Sub-sarcolemmal build-up of filamentous aggregates (white arrows) indicating further the impaired protein folding process as the major cause of the myopathic changes in this patient. Scale bar = 1 µm. The TEM studies were performed at the Institute of Neuropathology, Aachen.

Based on the subcellular localization information; ~95% of total altered proteins belong to the vital cellular components i.e. the nucleus (19), the cytoplasm (47) and the mitochondria (20) indicating a high degree of physiological disturbances in skeletal muscle of the index patient. The main cytosolic proteins i.e. actin and myosin that play an important role in the proper function of muscle cells such as contraction/relaxation and cytoskeletal organization were found to be upregulated in the index patient. Some of the well-known examples include: actin cytoplasmic 1 and 2 (*ACTB* and *ACTG1*), actin, alpha skeletal muscle (*ACTA1*); myosins 1, 4, and 13 (*MYH1*, *MYH4* and *MYH13*). Apart from these, other proteins which are involved in the movement e.g. tubulin beta chain (*TUBB*), tubulin alpha-4A chain (*TUBA4A*) and maintenance of the structural integrity of muscle fibers e.g. thymosin beta-4 and

beta10 (*TMSB4X* and *TMSB10*) also showed increased abundances in the index patient when compared to healthy controls, indicating a myopathic condition. These results are in accordance with the TEM findings (Fig. 4.11 A and B), which revealed electron-dense material and filamentous aggregates in the sarcoplasm (i.e. cytoplasm of muscle cells) of the muscle specimen of the index patient. Furthermore, increased levels of the cytoskeletal components indicate a possible formation of protein aggregates since upregulation of cytosolic molecular chaperones was found in the patient's data which include heat shock protein HSP 90-alpha (*HSP90AA1*), heat shock-related 70 kDa protein 2 (*HSPA2*), heat shock cognate 71 kDa protein (*HSPA8*), heat shock 70 kDa protein 1A and 1B (*HSPA1A* and *HSPA1B*). These chaperones usually operate in a concerted manner to modulate protein aggregation in the cytoplasm. Besides, kelch-like protein 41 (*KLHL41*) - a cytosolic protein involved in development and differentiation of skeletal muscle fibers, was among upregulated proteins. Notably, *KLHL41* has been associated with an autosomal recessive form of Nemaline myopathy (NEM9, OMIM: 615731) and one of its phenotypic manifestations is progressive distal muscle weakness starting at an early age, which was also observed in the index patient. In this regard, ryanodine receptor 1 (*RYR1*), a sarcoplasmic reticulum (i.e. sER in muscle cells) membrane bound Ca^{2+} ion channel also showed increased abundance in the index patient. Strikingly, mutations in *RYR1* have been linked to the central core disease (CCD, OMIM: 117000) - a skeletal muscle disorder. Moreover, both NEM9 and CCD are classified as congenital myopathies and the clinical hallmark of these disorders is the presence of rod-like structures in muscle cells [164].

Furthermore, upregulation of the mitochondrial ATP synthase subunit beta (*ATP5B*) that participates in the ATP production suggest an energy demand due to the cellular burden caused due to protein aggregation. Additionally, cytochrome c oxidase subunit 2 (*MT-CO2*) is involved in the respiratory chain in the mitochondria showed increased levels in the index patient, which is indicative of apoptosis. Recently, a novel mutation was identified on the *MT-CO2* that causes mitochondrial dysfunction leading to myopathy and neuromuscular disorders [165]. Histones are the basic proteins which reside in the nucleus and play a major role in gene regulation. In total, 9 different types of histones (out of 19 nucleus-residing proteins) were upregulated in the patient indicating an increased transcription regulation of proteins e.g. molecular chaperones, which might be involved in the ER quality control to cope with the cellular stress. Notably,

creatinine kinase m type (*CKM*), which is regarded as an indirect marker of muscle damage [166] was upregulated in the patient and this correlates with the biochemical results that detected moderately increased levels of *CKM*.

Among the downregulated proteins, decreased abundance of synaptopodin (*SYNPO*) - an actin binding protein, which is involved in the formation of telencephalic neurons, can be associated to the neuromuscular abnormalities presented by the index patient. *FKBP3* - a member of the peptidyl prolyl cis/trans isomerases (PPIases) family, which catalyzes the cis-trans isomerization of proline peptide (Xaa-Pro) bonds in oligopeptides and also accelerates the protein folding process [167], was downregulated indicating protein aggregation. In addition, eukaryotic translation initiation factor 4B (*EIF4B*) - an mRNA binding protein, which is involved in the protein biosynthesis, showed low levels suggesting a possible translational attenuation as a consequence of the UPR pathway initiation. Further evidence of the UPR activation is revealed by the upregulation of ubiquitin-conjugating enzyme E2 N (*UBE2N*) and polyubiquitin-B (*UBB*), both involved in the ERAD-mediated protein ubiquitination pathway.

4.5 Proteomic profiling of SIL1-depleted HEK293 cell line

From the above results, it can be deduced that the loss of functional SIL1 may lead to cellular perturbations owing to dysfunctional BiP-controlled protein folding in the ER-lumen. However, due to the absence of appropriate *in vitro* cell models, the precise cellular pathophysiological mechanisms leading to neuromuscular degeneration in MSS are still unclear [168]. For this, a comparative proteomics analysis was performed using SIL1-depleted human embryonic kidney 293 (HEK293) cells to better understand the structural changes of the ER that closely mimic pathological alterations in MSS due to the absence of SIL1. The two SIL1-depleted cell clones (ΔSIL1_1 and ΔSIL1_2) and one scrambled (Scr) transfected cell clone (as a negative control) were prepared using the FASP protocol (see Section 3.2.6). Next the samples were analyzed by label-free quantitative proteomics. Precursor ion area-based quantification (see Section 3.2.15.2.1) led to the quantification of 2,819 proteins (≥ 2 unique peptides, 1% FDR) of which 459 (16%) showed altered levels (set criteria: ANOVA p-value ≤ 0.05, an average ΔSIL1/Scr ratio ≤ 0.667 or ≥ 1.6 for down-/upregulation). Of these, 141 (31%) were up- and 318 (69%) were

downregulated in HEK293-ΔSIL1 cell line, respectively. Details of altered proteins are given in the Appendix 10.9.

By using the gene ontology (GO) [169] analysis in conjunction with the STRING database [170], differentially regulated proteins (or genes) were enriched under the domain "biological process" term to better understand the effects due to SIL1-loss. The "GO term analysis" analyzes datasets for the overrepresentation of specific biological terms or functions by making use of the GO classification, in which genes are assigned to a set of predefined terms depending on their functional characteristics. The GO analysis results of the altered proteins revealed that they are involved in important cellular pathways and are described below in detail. However, due to a high number of differentially regulated proteins in this dataset, only their gene names are used as shown in the Fig. 4.12.

Organization and proper function of cytoskeleton relies on the number and spatial arrangement of membrane proteins. They render binding sites e.g. to actin filaments and microtubules, and are possible substrates of SIL1-BiP machinery. In HEK293-ΔSIL1 cell line, altered levels of the plasma membrane-associated proteins e.g. *CKAP4*, *FERMT2*, *RDX*, *TLN1* and *VCL* were identified. In addition, increased levels of *PFN1*, *CFL1*, *PDXP* and *DSTN* (an actin-depolymerizing protein) and decreased abundance of *GSN* were found. *GSN* is Ca^{2+}- regulated, actin-modulating protein that can promote assembly of monomers into filaments (nucleation) as well as repair filaments already formed. Downregulation of *GSN* could indicate an altered ER homeostasis or suggest an alternate mechanism to counteract cytoskeletal disturbance [168]. Increased abundances of *TBCA*, *TBCE* and *TBCB*, which provide attachment of cytoskeletal filaments to plasma membrane proteins, indeed support the latter assumption. Besides, downregulation of cytoskeleton proteins such as, *EPPK1*, *TUBGCP3*, *MYO1C*, *MYH10*, *MAP1B*, *NEFL*, *NEFM* and *VIM* were also detected. Notably, *VIM* plays a role in the progression of cataracts [OMIM: 116300] - one of the clinical hallmarks of MSS [168]. Drebrin-like protein (*DBNL*) interacts with Dynamin (*DNM2*), which plays an important role in vesicular trafficking processes and also involved in various neuromuscular disorders e.g. Charcot-Marie-Tooth disease [OMIM: 606482]. Increased abundances of these two candidates indicate an altered protein trafficking between the ER, the Golgi and the plasma membrane.

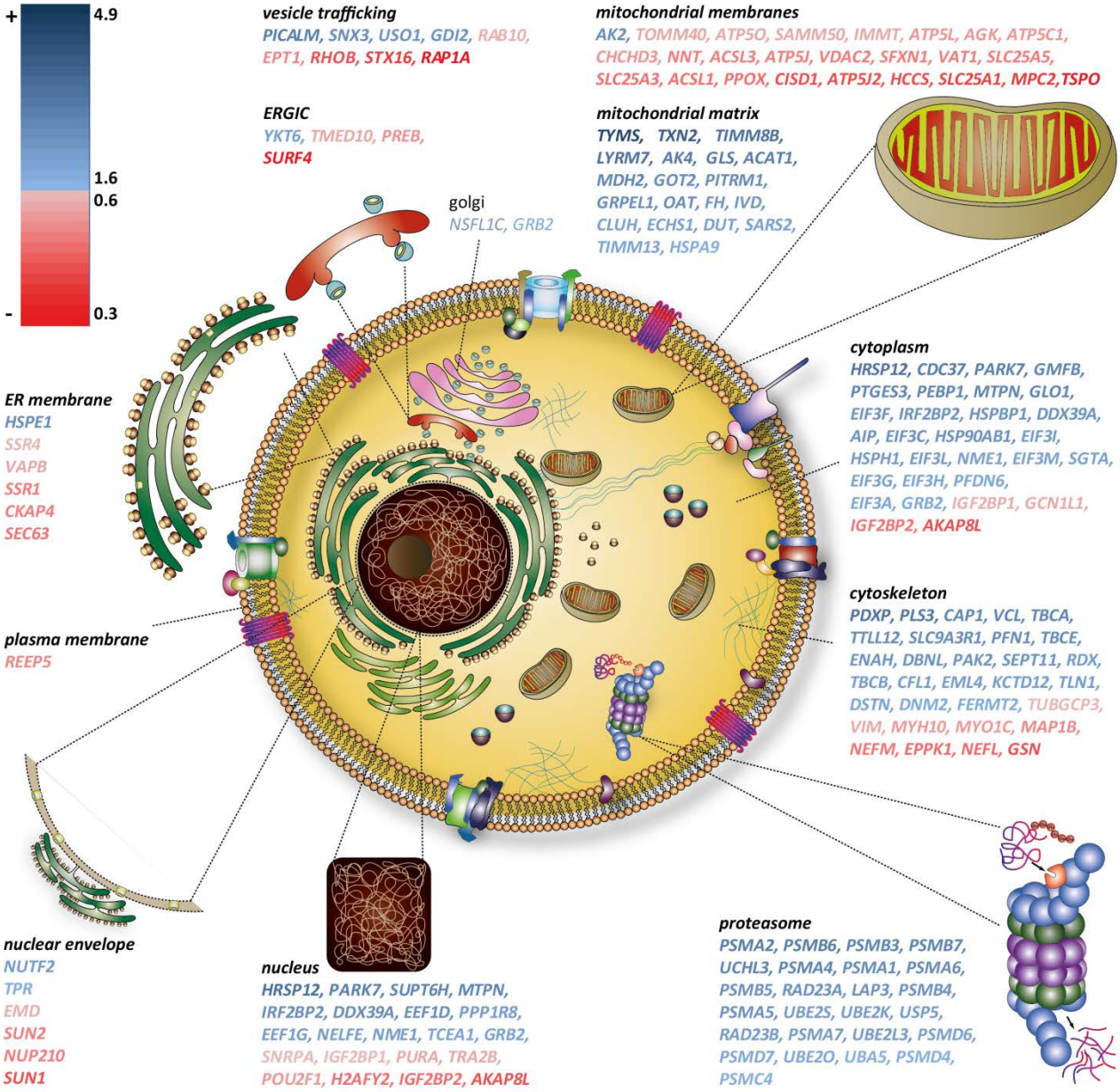


Figure 4.12: Graphical representation of different subcellular compartments and the respective resident proteins that showed significant regulation (set criteria: ΔSIL1/Scr ratio of ≥ 1.6-fold, log-2 scale; ANOVA p-value ≤ 0.05) determined by label-free quantitative MS data as described in detail in the section 3.2.15.2.1. The gene name color coding represents increased (blue) and decreased (red) protein abundances. Figure adapted from [168].

Further evidence of disturbed protein transport system is shown by downregulation of *RAB10*, *RHOB*, *RAP1A*, *EPT1*, *SURF4*, *STX16*, *TMED10* and *VAPB*, and upregulation of vesicular transport factors, such as *USO1*, *PICALM*, *SNX3* and *YKT6* (Fig. 4.12). Moreover, this data is in line with the TEM data that revealed proliferated Golgi and vesicular structures (Fig. 4.13 B). Together, altered levels of cytoskeletal and vesicular transport proteins might be due to dysfunctional

trafficking mechanism caused by SIL1-loss leading to an abnormal levels of the plasma membrane proteins including decreased levels of *REEP5* (Fig. 4.12) [168].

Mitochondrial damage is oftentimes represented by its aberrant shape as observed in the MSS-LCs and *woozy* mice tissues. Despite SIL1 being an ER-resident protein, its association with the mitochondrial dysfunction could be explained by the following evidences: (i) mitochondria-ER associated membranes (MAMs) are specific domains that enable multiple adhesion sites between both organelles, (ii) mitochondria-ER interaction promotes mitochondrial fission, and (iii) actin polymerization, which is affected by SIL1-loss (see above) and this in turn modulates mitochondrial fission [168]. Thus, cytoskeletal and mitochondrial disturbances can be ascribed to SIL1-depletion. The latter assumption is therefore supported by decreased abundances of several mitochondrial membrane and membrane binding proteins e.g. *ACSL1*, *ACSL3*, *AGK*, *ATP5C1*, *ATP5J*, *ATP5J2*, *ATP5L*, *ATP5O*, *CISD1*, *HCCS*, *IMMT*, *MPC2*, *NNT*, *PPOX*, *SAMM50*, *SFXN1*, *SLC25A1*, *SLC25A3*, *SLC25A5*, *TOMM40*, *TSPO*, *VAT1*, *VDAC2* (Fig. 4.12). Furthermore, downregulation of *CHCHD3*, which plays crucial roles in the maintenance of different mitochondrial membranes including cristae, indicates the impact of SIL1-loss on this subcellular organelle. These findings correlate with mitochondrial alterations detected on the TEM level (Fig. 4.13 B, C) [168]. In addition, HEK293 ΔSIL1 cells showed an increased abundance of a subset of mitochondrial matrix/luminal proteins that might counteract mitophagy (i.e. selective degradation of mitochondria by autophagy). These proteins take part in proper cytoplasmic mitochondrial distribution (*CLUH*), lipid (*ECHS1*, *HADH*) and ketone body (*ACAT1*) metabolism, tricarboxylic acid cycle (*FH*, *MDH2*), cell redox homeostasis (*TXN2*), amino acid metabolism (*GOT2*, *IVD*, *OAT*), nucleotide (*AK2*, *AK4*, *DUT*, *TYMS*) and protein biosynthesis (*SARS2*), protein folding (*GRPEL1*, *HSPA9*, *LYRM7*, *TIMM8B* and *TIMM13*) and proteolysis (*PITRM1*) (Fig. 4.12) [168]. Notably, these results are in accordance with the observed mitochondrial alterations in MSS [35] for which SIL1 was described as a modulator [45b] to mitochondrial dysfunction [171].

The alterations of nuclear envelope architecture are hallmarks of SIL1/Sil1 chaperonopathies in skeletal muscle fibers in man and mouse [35]. The TEM studies on ΔSIL1 HEK293 cells showed abnormal nuclear morphology including the nuclear envelope and the nuclear membrane (Fig. 4.13 B). The occurrence of these disturbances can be attributed to differential regulation of nuclear envelope related proteins and the nuclear pore complex proteins i.e. *TPR*, *NUP210*,

SUN1/2, *NUTF2*, and *EMD* including *VAPB* (Fig. 4.12). Other TEM findings include elongated ER particularly in the perinuclear region (Fig. 4.13 B, C), abnormal lysosomes and autophagic vacuoles (Fig. 4.13 D). To support this, the proteomics data revealed altered levels of ~28% ER-membrane proteins, implying that these are potential targets of the SIL1-BiP protein folding complex [168]. Whereas, increased levels of ubiquitin hydrolases e.g. *UCHL3*, *USP5* and proteases e.g. *LAP3*, *CTSD* indicate activation of lysosome-associated autophagy. Notably, *CTSD* is involved in the pathogenesis of several disorders including the Alzheimer's disease [168].

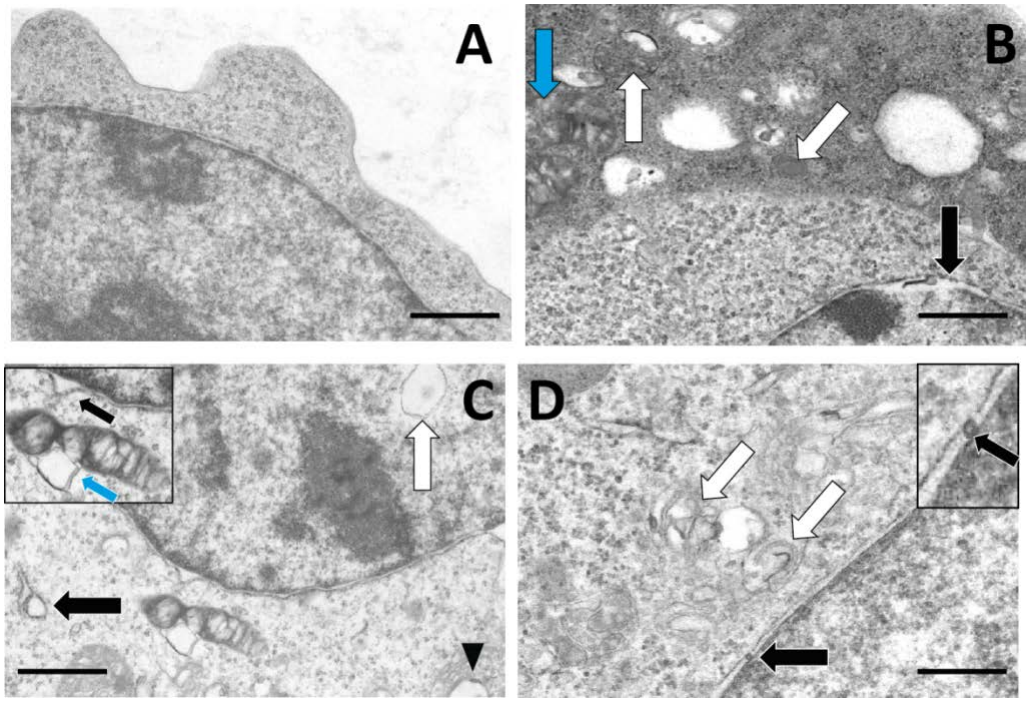


Figure 4.13: TEM data in (**A**) scrambled transfected HEK293 cells showing normal organelle morphology. Scale bar=1.0 µm. Whereas, ΔSIL1 HEK293 cells showed (**B**) elongated ER structures, abnormal mitochondria (blue arrow), lysosomes (white arrows), and altered nuclear morphology (black arrow). Scale bar=0.25 µm. Further, (**C**) widened rER (large black arrow) and altered nuclear envelope (black arrow in inset) were also detected. Mitochondrion damage is indicated with an osmiophilic membrane (blue arrow in inset) and prominent Golgi-cisterns (black arrowhead). Scale bar=1.0 µm. Lastly, (**D**) Lysosomal/autophagic material (white arrows) in the perinuclear cytoplasm and an osmiophilic inclusion in the nuclear envelope (black arrows) were found. Scale bar=0.5 µm. Figure adapted from [168].

Based on these subcellular organelle perturbations, it was evident that the ER-stress induced UPR pathway has been triggered. Therefore, altered regulation of the UPR-related proteins was verified in the label-free analysis and further validated by the targeted PRM-based assay [116] (see Sections 3.2.13 and 3.2.15.3). These results suggest that the overall protein synthesis is reduced owing to (i) a decrease of transcription assisting factors: *POU2F1*, *PURA*, *SNRPA*, *TRA2B*, (ii) an

increase of the repressors: *HRSP12* and *PPP1R8*, (iii) a decrease of *IGF2BP1* and 2 that are involved in the transport and transient storage of mRNA and of the translational activator - *GCN1L1* and (iv) reduced levels of the translocon proteins: *SEC63*, *SSR1*, *SSR4* and *MAP1B* (Figs. 4.12 and 4.14 A). Additionally, the eukaryotic translation initiation factor 3 (*EIF3*) complex is required for disassembly of ribosomes, upregulation of its subunits A, C, F, G, H, I, L and M (Fig. 4.12) is consistent with the translational inhibition. Besides, increased abundances of certain transcription and translation factors, such as *PAF1*, *TCEA1*, *EEF1G*, and *EEF1D* (Fig. 4.12) are a hallmark of the UPR pathway and correspond to the induction of stress related factors [172]. Further, targeted-PRM assay results showed increased levels of the UPR-pathway assisting factors i.e. BiP, GRP75, GRP94, GRP170, *PDI* and *CALR* (Figs. 4.14 A). This data is in good correlation with the WB analysis (4.14 B).

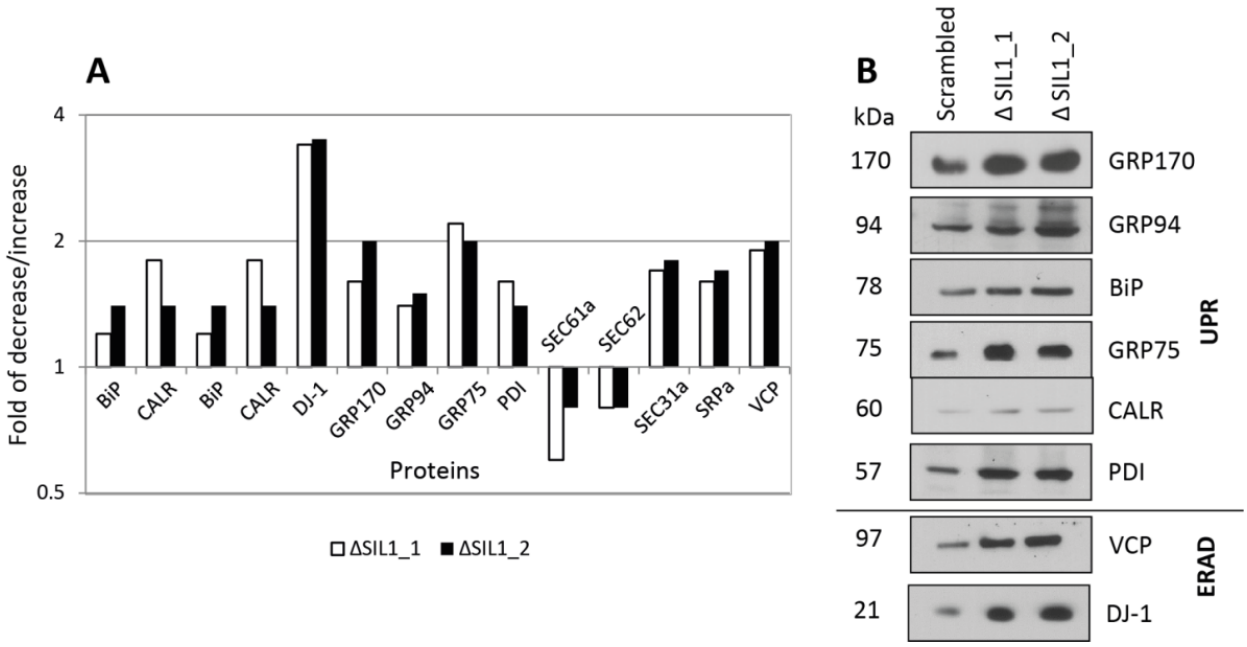


Figure 4.14: Biochemical findings in scrambled transfected and SIL1-depleted HEK293 cells. **(A)** PRM-based targeted-MS analysis showed increase of UPR assisting factors including: GRP75 (~120%), BiP (~40%), GRP94 (~50%), GRP170 (~100%), PDI (~60%), CALR (~80%), SEC31a (~80%), and SRP-α (~70%). Also in agreement with UPR activation, translocon proteins SEC61a (~40 %) and SEC62 (~20%) were decreased. Moreover, elevated levels of VCP (~100%) and DJ-1 (~250%) were indicative for ERAD activation. **(B)** The targeted-MS data was supported by the WB studies that reveal the activation of the UPR pathway due to the ER stress and protein clearance in SIL1-depleted compared to scrambled transfected HEK293 cells. Figure adapted from [168].

4.6 Characterization of human myoblastic cell line - RCMH

As myopathy is one of the clinical hallmarks of MSS and due to limited availability of the muscle biopsies from these patients; a human myoblastic cell line e.g. RCMH would be an ideal choice for (i) studying physiological processes, (ii) discovering pathological processes and (iii) developing novel cell-based therapies for muscular disorders [173]. High-pH (6.0) RP fractionation of the tryptic peptides generated from the RCMH cells followed by LC-MS analysis and database searches of all 16 MS raw files led to the identification of 6,244 proteins with a FDR of ≤ 1% on the PSM, peptide, and protein level. Notably, the subsequent calculation of the NSAF values (see Section 3.2.15.2.2) of these proteins yielded a range of more than four orders of magnitude between the most abundant protein actin (*ACTG1*) and the ryanodine receptor 3 (*RYR3*). According to the UniProtKB: 2,132 proteins are localized in the cytoplasm (398 cytoskeletal), 1,957 in the nucleus, 500 in mitochondria, 355 in the ER, 264 in the Golgi and 69 in lysosomes, whereas 169 proteins are secreted. The obtained NSAF values were then compared with two other cell types with well-characterized proteomes namely, HeLa and U2OS in order to identify an up-/downregulation of certain pathways and biological functions in RCMH cells and to deduce a more functional picture. For this, MS raw data from Guo et al., (HeLa) [174] and Beck et al., (U2OS) [175] were downloaded from PRIDE [176] repository and re-searched with the same software and settings as used for the RCMH data. This led to the identification of 6,304 proteins in HeLa and 7,158 proteins in U2OS compared to the 6,244 proteins in RCMH. Notably, these cell lines were selected owing to the still sparse availability of comprehensive high quality datasets for human cell lines.

To classify the composition of these three apparently different proteomes, all the identified proteins in each dataset were analyzed for an enrichment of GO terms using the Ontologizer software [177] compared to the human UniProt database (downloaded on 11th of December 2013, containing 20,273 target sequences). Indeed, this analysis did not show major differences between RCMH and other cell types. This can be hypothesized as a possible consequence of the undifferentiated stage of the RCMH cell line in the culture conditions utilized, which favor proliferation rather than fusion and myotube formation *in vitro* [178]. Moreover, this demonstrates a limitation of the general way the GO enrichment analyses are performed, as

these merely take into account the presence of proteins in lists rather than the actual expression levels, which are present in the used quantitative proteomics data. Consequently, as long as similar numbers of proteins for a given GO term are detected in two different cell types, they produce similar results, even if the respective proteins are considerably higher expressed in one of the two cell types. Therefore, only a subset of GO terms with respect to the origin of these cells that play a role in muscle function or are involved in myopathic disorders were considered and the NSAF values of the corresponding proteins were extracted in each sample set i.e. cytoskeleton (GO:0005856), proteasome complex (GO:0000502), striated muscle (combination of adaptation (GO:0014888), atrophy (GO:0014891), cell development (GO:0055002), nuclear part (GO:0044428), plasma membrane (GO:0005886), mitochondrion (GO:0005739), regulation of response to stress (GO:0080134), muscle cell development (GO:0055001), metabolic process (GO:0008152), endoplasmic reticulum (GO:0005783), Golgi apparatus (GO:0005794), neuromuscular process (GO:0050905), and autophagy (GO:0006914). Next, for each GO term and sample set, the number of corresponding proteins was determined and the sums of their NSAF values were calculated as well as the average NSAF (NSAF sum divided by number of proteins), which reflects the relative expression of the particular GO term for a representative comparison of the different proteomes. Thus, the differences in the expression patterns of RCMH, HeLa and U2OS became apparent for instance, proteins corresponding to striated muscle are clearly more abundant in RCMH than in U2OS and HeLa and the same holds true for proteins involved in muscle cell development. Furthermore, the expression levels of proteins related to the neuromuscular process are marginally higher in RCMH compared to HeLa, but remained nearly the same in U2OS (see Table 4.3).

Identification of cytoskeletal components responsible for contraction can be considered as a rationale in the discovery of a suitable cell model for *in vitro* muscle investigations. The levels of actin and myosin - the chief structural proteins, were higher in RCMH cells compared to HeLa and U2OS. For instance, the NSAF value (normalized to *GAPDH* NSAF value) of actin, cytoplasmic 1 (*ACTB*) was 3.6-fold and 2.1-fold higher in RCMH compared to HeLa and U2OS, respectively. Whereas, the abundances of different unconventional myosin light and heavy chains (including 12 different unconventional myosins, 3 tropomyosins and myosin phosphatase Rho interacting protein) were more in RCMH compared to HeLa (~5-fold higher) and U2OS (~1.5-fold higher).

Table 4.3: Comparison of GO term expression levels between RCMH, HeLa and U2OS cells based on both NSAF sum and NSAF average.

GO Term	sum RCMH/HeLa	average RCMH/HeLa	sum RCMH/U2OS	average RCMH/U2OS
cytoskeleton	0.96	0.97	1.14	1.29
proteasome complex	0.85	0.86	0.97	1.01
striated muscle	1.54	1.21	1.47	1.54
nuclear part	0.97	0.99	0.99	1.12
plasma membrane	1.03	0.98	1.10	1.23
mitochondrion	0.77	0.91	0.92	1.11
regulation of response to stress	1.02	0.75	1.06	0.83
muscle cell development	1.95	1.61	1.70	1.58
metabolic process	0.91	0.92	0.98	1.10
endoplasmic reticulum	0.98	1.00	1.14	1.29
Golgi apparatus	1.07	1.05	1.04	1.06
neuromuscular process	1.24	1.50	0.82	0.95
autophagy	0.95	0.96	0.61	0.63

Besides actin and myosin, expression of other cytoskeletal components e.g. desmin (*DES*) and titin (*TTN*) was also identified in *in vitro* muscle studies [179]. In RCMH cell line, desmin is among the 100 most abundant proteins, which was not identified in other two cell lines. Furthermore, proteins that are involved in (i) maintaining stability of the cytoskeleton, such as α- and β-spectrins (~4.4 and 6.4-fold higher than in HeLa, same in U2OS), (ii) anchoring of other cytosolic proteins e.g. filamins (actin-binding proteins) and the respective assembly promoting proteins i.e. α-, β- and γ-adducin (~9.2-fold higher than in HeLa, same in U2OS) including dystrophin (*DMD*) (~8-fold higher than in HeLa, same in U2OS) were detected in RCMH. These proteins are known to modulate mechanotransducer action - an important process that helps muscle fibers to cope with mechanical stress [180]. In mammalian muscle cell, *DMD* connects cytoskeletal actin via the dystroglycan complex (DGC) to laminin (*LAMB1*) localized in the extracellular matrix. Apart from dystrophin, other components of DGC e.g. dystroglycan (*DAG1*; expression highest in RCMH), β-sarcoglycan (*SGCB*; ~9.2-fold higher than in HeLa, not detected in U2OS), Δ-sarcoglycan (*SGCD*; only detected in RCMH), α-1-syntrophin (*SNTA1*; ~16-fold higher than in HeLa, not detected in U2OS), β-1-syntrophin (*SNTB1*; only detected in RCMH), β-2-syntrophin (*SNTB2*; highest expression in RCMH), and dystrobrevin beta (*DTNB*; 33-fold higher in RCMH than in HeLa, not detected in U2OS). Notably, some of these proteins have been reported to be involved in various types of myopathies e.g. mutations in *DMD* cause Duchenne muscular

dystrophy [OMIM: 310200] - a muscle wasting disease mainly affecting young adult males [181]. Additionally, RCMH proteome also includes *NOS3* (not found in HeLa and U2OS), the endothelial isoform of the nitric oxide synthase as well as DGC assisting factors like ankyrin-3 (*ANK3*), which are required for costamere localization of dystrophin and *DAG1* [182].

As seen in *woozy* mice and human index patient muscle datasets, the dysfunctional or damaged mitochondria can severely perturb metabolic processes leading to muscular disorders. Therefore, expression of the known mitochondrial proteins including succinate dehydrogenase subunits e.g. *SDHA*, *SDHB*, *SDHC*, and of *SDHAF2* as well as of the cytochrome c oxidase subunits e.g. *COA3*, *COX5A*, *COX6B1*, *COX6C*, *COX7A2L*, *COX17* and *TACO1* in RCMH suggests that these cells are a suitable model for investigating metabolic processes in muscle. Interestingly, the expression levels of all succinate dehydrogenase subunits were higher in RCMH compared to HeLa, but remained the same in U2OS. The same trend was observed in the case of cytochrome c oxidase subunits. It is evident that the ER-stress caused due to the alterations in the protein folding machinery invariably triggers the cellular defense mechanisms, particularly the UPR and ERAD. Expression of proteins which belong to these pathways e.g. *ATF6*, *HSPA5* (BiP), *CALR*, *CANX*, *DNAJB6*, *EDEM2* and *3*, *EIF2AK3* (PERK), *ERN1* (IRE1), *HSPA1A*, *HYOU1* (GRP170), *P4HB*, *SIL1* and *UGGT* (see Table 4.4) in RCMH suggests that this cell line can be used to investigate muscle-related pathophysiological processes.

Table 4.4: NSAF value-based comparison of proteins that are part of the UPR, ERAD pathways and nuclear envelope integrity (NEI) between RCMH, HeLa and U2OS cells; (ND = not detected).

Protein	Gene	NSAF RCMH/HeLa	NSAF RCMH/U2OS	Function
Cyclic AMP-dependent transcription factor ATF-6 alpha	*ATF6*	ND	ND	UPR
78 kDa glucose-regulated protein	*HSPA5*	1.4	1.8	UPR, ERAD, NEI
Calreticulin	*CALR*	1.5	1.4	UPR
Calnexin	*CANX*	1.3	1.5	UPR, NEI
DnaJ homolog subfamily B member 6	*DNAJB6*	6.9	1.3	ERAD
ER degradation-enhancing alpha-mannosidase-like protein 2	*EDEM2*	ND	1.3	UPR
ER degradation-enhancing alpha-mannosidase-like protein 3	*EDEM3*	1.9	2.2	UPR
Eukaryotic translation initiation factor 2-alpha kinase 3	*EIF2AK3*	24.5	ND	UPR
Serine/threonine-protein kinase/endoribonuclease IRE1	*ERN1*	ND	4.6	UPR
Heat shock 70 kDa protein 1A/1B	*HSPA1A*	0.9	0.5	ERAD
Hypoxia up-regulated protein 1	*HYOU1*	1.7	0.9	UPR
Protein disulfide-isomerase	*P4HB*	1.1	1.6	UPR
Nucleotide exchange factor SIL1	*SIL1*	4.1	2.6	UPR, ERAD, NEI
UDP-glucose:glycoprotein glucosyltransferase 1	*UGGT*	1.7	0.5	UPR

Notably, all three major transducers of the UPR pathway i.e. *ATF6*, *EIF2AK3* and *ERN1* were identified in RCMH, but *ATF6* was not detected in other two cell lines. Whereas, *EIF2AK3* was not detected in U2OS and *ERN1* was not identified in HeLa. Another aspect related to muscle physiology is the excitation-contraction-coupling (EC-coupling), whereby the electrical impulse (excitation) from the neuron is converted into a mechanical response (contraction) at the neuromuscular junction. In skeletal muscle, EC-coupling relies on a direct connection between the ryanodine receptor (*RYR*), a sarcoplasmic reticulum Ca^{2+} release channel and dihydropyridine receptors (DHPRs), acting as voltage-gated L-type Ca^{2+} channels [183]. DHPRs are located on the sarcolemma (i.e. cell membrane of muscle cells), which also comprises the transverse tubules (i.e. invaginations of the sarcolemma). When an action potential (electrical impulse) is generated and propagated across the sarcolemma, it depolarizes the sarcolemma resulting in an increase of cytosolic Ca^{2+} concentration, which leads to muscle contraction in a Ca^{2+} dependent manner. In RCMH cells, expression of *RYR* and DHPR (*QDPR*) has been reported earlier [184]. In line with these findings, expression of Ca^{2+} associated proteins, such as *ATP2A2*, *ATP2B1*, *ATP2B4*, *CACNA1S*, *ESYT1*, *ESYT2*, *KCNMA1*, *CAMK2D*, *CAMK2G* and *CHERP* were detected in RCMH. Besides, *CALM1*, *CALU*, and *STIM1* were also identified. Notably, accumulation of the latter proteins was linked with muscle fiber degeneration in neurogenic muscular atrophies, in which EC-coupling is altered due to neuronal dysfunction [185]. Moreover, the identified α- and β- chains of spectrin (*SPTAN1*, *SPTBN1*) interact with *CALM* in a Ca^{2+} dependent manner and play an important for muscle contraction [186].

Additionally, morphological studies on RCMH cells (cultured for 24 hours) were performed using scanning electron microscopy (SEM) and TEM (data not shown). The SEM studies revealed spread cells that exhibit cytoplasmic extensions. These microscopic protrusions of the plasma membrane not only increase the cell surface, but are also responsible for cellular communication and mechanotransduction. TEM studies confirmed the presence of cytoplasmic extensions and revealed prominent ribosome dense sarco-/endoplasmic reticulum networks. This morphological finding is in line with the expression of > 300 sarcoplasmic reticulum-associated proteins identified by proteomics analysis [173].

5 Discussion and conclusions

In this work, proteomic investigation of the Marinesco-Sjögren Syndrome (MSS) by MS-based quantitative approaches was performed to better understand the pathophysiology of this disorder. For this, unaffected (i.e. five different MSS cases; SIL1 non-vulnerable) and affected (i.e. *woozy* mouse; SIL1 vulnerable) tissues were studied (using iTRAQ) to gain insights into the role of SIL1 in BiP-mediated protein folding process in the ER. Additionally, characterization of *in vitro* models (i.e. HEK293-ΔSIL1 and RCMH) was done to evaluate their suitability for studying neurodegenerative and neuromuscular disorders. Furthermore, sample preparation induced artificial modification (i.e. carbamylation of proteins/peptides) that might hinder labeling procedures (e.g. iTRAQ) and affect quantification accuracy, was studied in detail.

5.1 *In vitro* protein carbamylation - a potential unwanted artefact

Initially, urea was added to the digestion buffer (final concentration of 1.0 M, overnight incubation at 37°C) for processing the human fibroblasts (MSS2 study). Notably, the share of carbamylated PSMs identified (at 1% FDR) in the respective iTRAQ 4-plex dataset was found to be < 1% (data not shown). Although the number of carbamylated PSMs was relatively low, it gave a hint that there might be a possibility to introduce this artefact under certain conditions involving urea during sample preparation. Hence, as urea is the most frequently used organic compound in majority of the sample preparation protocols (e.g. FASP) for both qualitative and quantitative proteomics analyses, a systematic study was performed using a pool of peptide mixtures generated from five model proteins. The degree of carbamylation reaction at the N-termini and epsilon amino group of Lys residues varied depending on the concentration of urea, exposure time with urea and heat. For instance, aqueous alkaline buffer solutions containing proteins or peptides together with urea at high concentrations (8.0 M) when heated at temperatures > 37°C for shorter time period (i.e. 30 min) resulted in ~13% of carbamylation. Whereas, at relatively low concentration i.e. 2.0 M, but longer incubation times at 37°C lead to almost 23% of carbamylation of N-termini. However, when a complex cell lysate (fibroblasts) was subjected to enzymatic proteolysis in presence of 2.0 M urea and 22 h of incubation at 37°C [120], carbamylation of < 0.3% of the identified peptides was detected, which is indeed below the

applied FDR of ≤ 1% and thus not significant. Moreover, there were no differences detected between the urea-based and urea-free samples. While this is in contrast to the results above, it can be most likely attributed to two major reasons. Firstly, undersampling due to performing the MS analysis in a DDA mode of complex proteomes [187] will reduce the number of carbamylated peptides, since for each of the 100,000s of peptides in the complex digest, only a negligible number of the present copies are actually carbamylated. Secondly, since the N-termini have a higher reactivity for carbamylation than Lys residues - this can cause an overestimation of carbamylated peptides in the already pre-digested reference sample in comparison to the fibroblast digest, in which new N-termini still have to be generated during the enzymatic digestion [122].

Nevertheless, the peptide mixtures data demonstrate that the usage of urea results in *in vitro* carbamylation with N-termini being the most susceptible followed by Lys residues, whereas carbamylation of Arg residues is spared under commonly used sample processing conditions. Therefore, for the experiments following the MSS fibroblasts iTRAQ 4-plex study, urea-containing buffers were freshly prepared and their usage was restricted only to eliminate the SDS from the samples using the FASP protocol (see Section 3.2.6). The remaining steps, such as carbamidomethylation (heating samples at 56°C for 30 min, see Section 3.2.5) and proteolytic digestion (overnight incubation at 37°C), were performed in the absence of urea. For the latter step, GuHCl was used as an alternate chaotrope to facilitate trypsin digestion.

5.2 Optimized sample preparation for quantification of SIL1

Despite its important role in the protein folding process, SIL1 is apparently present in relatively low abundance compared to GRP170 and BiP in different eukaryotic cell lines/types (Fig. 5.1). Therefore, in order to relatively quantify such a low-stoichiometric protein, robust sample preparation workflows must be employed that can provide reliable results. In recent years, several sample preparation techniques emerged with the advent of MS-based proteomics that primarily focused on removing chemicals that interfere with proteolysis and downstream LC-MS analysis. These mainly include in gel, in solution and filer-based (e.g. FASP) digestion protocols.

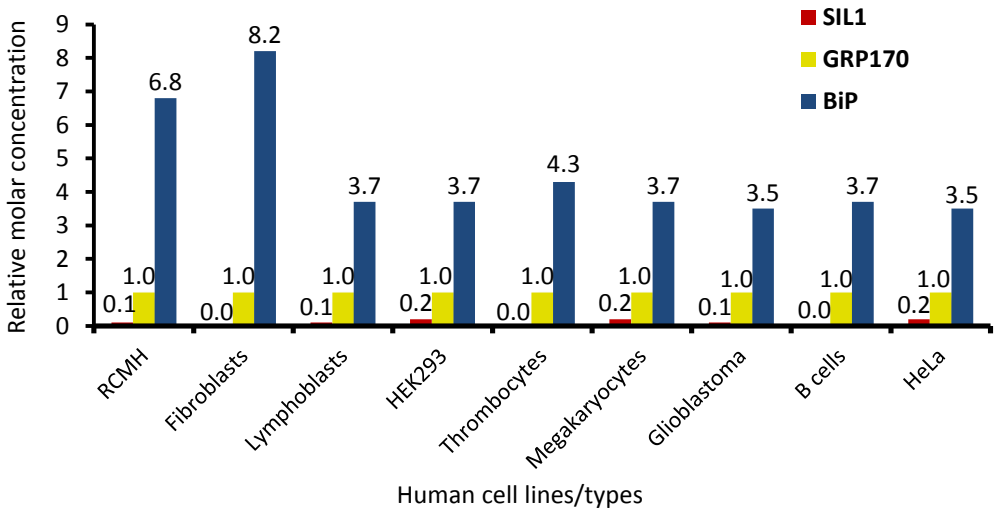


Figure 5.1: Comparison between the relative molar concentration of SIL1, GRP170 and BiP in different human (i) cell lines i.e. RCMH, fibroblasts, LCs, HEK293, glioblastoma (A-172), HeLa and (ii) cell types i.e. thrombocytes (platelets), megakaryocytes, B cells (B lymphocytes). The molar concentrations of SIL1 and BiP are normalized with respect with GRP170. The numbers are based on the NSAF values that were generated from the label-free MS data of each cell line/type.

Among those, the FASP technique has grown rapid popularity owing to its effectiveness in removing strong detergents (e.g. SDS) and tolerant to high concentrations of chaotropic agents (e.g. 8.0 M urea), which are frequently used to solubilize mainly membrane proteins. Moreover, FASP allows ease in sample handling and yields better proteome sequence coverages [120]. Furthermore, iTRAQ labeling (8-plex) indeed improved the sensitivity of the analysis (owing to multiplexing) and in conjunction with high pH-RP fractionation (Fig. 5.2) provided relative quantification of 4,858 proteins with ≥ 2 unique peptides in the MSS-LCs dataset. Compared to the detection and quantification rates of other proteomics studies using LCs, this is the most comprehensive proteome profile of human LCs achieved so far [188]. Besides, chemical labeling (4- and 8-plex) and offline fractionation strategies (OFFGEL-IEF and RP at pH 6.0) followed by LC-MS analysis enabled relative quantification of the low-abundant SIL1 in MSS-fibroblasts and MSS-LCs datasets (Figs. 4.4 and 4.6 A). However, as shown in the Figure 5.1, the relative molar ratio of other NEF to BiP i.e. GRP170 is nearly 10-fold higher, which in fact is also constant across all cell lines. Therefore, even if loss of SIL1 leads to increase in a NEF that is 10 times more abundant, the current accuracy of the quantitation strategies (e.g. iTRAQ) might be insufficient to verify this with high confidence. To achieve this, AQUA (absolute quantification) [189] strategy can be employed wherein, stable isotope labeled (SIL) peptides as internal standards are used

i.e. adding a known amount of SIL counterpart of an endogenous peptide prior to LC-MS analysis.

Furthermore, AQUA can also be combined with targeted-MS methods e.g. PRM for sensitive and accurate quantification of proteins [100] including the low abundant ones such as SIL1.

Whereas in *woozy* tissues i.e. cerebellum and skeletal muscles, Sil1 could not be not detected. Despite using the same workflow, the total number of quantified proteins in these two datasets (3,580 and 2,055, respectively) was relatively low compared to the primary cell lines data (4,858 proteins in MSS-LCs). The obvious reason for this difference is that the tissues are typically composed of multiple cell types and tend to be more complex with wide dynamic range of protein concentrations as opposed to the cell lines that comprise predominantly a single cell population. Therefore, direct analysis of crude tissue samples is complicated by the presence of abundant proteins which usually mask the low abundant candidates. This holds true particularly in *woozy* skeletal muscle data in which actin and myosin and their isoforms account for nearly one-third (34%) of quantified proteins. In order to

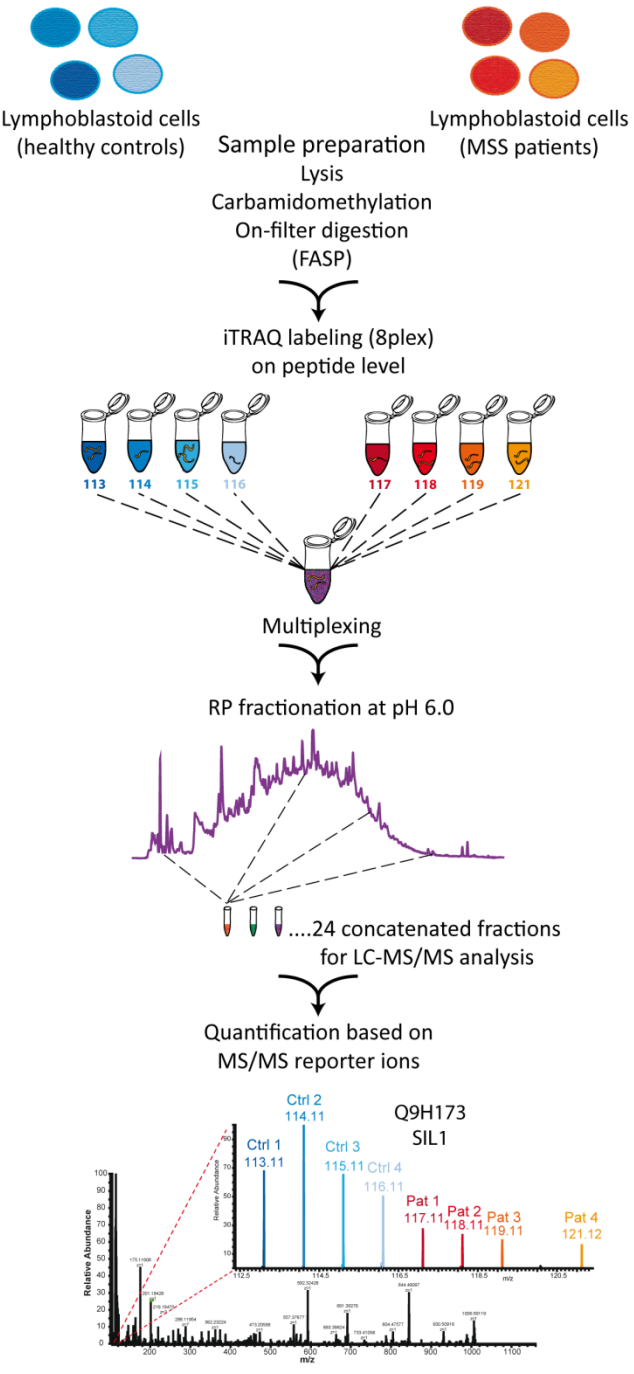
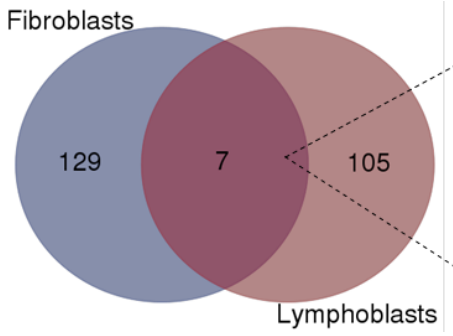


Figure 5.2: Bottom-up proteomics workflow involving FASP, iTRAQ 8-plex labeling, high-/low pH RP-RP chromatography and MS analysis. Fragment ion spectra of a SIL1 peptide depicting the reporter ions indicative of its relative abundances across eight different samples.

minimize the tissue complexity, protocols that are based on the subcellular organelle fractionation were developed to isolate for instance: nuclear, cytosolic, mitochondrial, ER compartments [190]. This approach in combination with chemical labeling (e.g. iTRAQ) has been reported to not only increase the proteome coverage of tissues, but also to quantify low abundant proteins that are located in specific organelles [191].

5.3 SIL1 loss - impact on unaffected cell types and compensatory mechanisms

Since *SIL1*/*Sil1* was identified as the pathogenic factor for MSS and *woozy* mouse phenotype [34c, 41b, 42], several studies showed that mutations of this gene cause neurodegeneration and myopathic changes in man and *woozy* mouse [41a, 47, 192]. However, the TEM findings (Fig. 4.3) revealed morphological alterations due to loss of SIL1, suggesting subclinical vulnerability in clinically unaffected tissues. Therefore, proteomics was used as a tool to verify these changes in the SIL1 non-vulnerable cell types (i.e. fibroblasts and LCs) and to understand the unknown rescue mechanisms that might be responsible for their survival. An iTRAQ-based quantification strategy revealed differential regulation of ~4.5% and ~2.3% of the total quantified proteins (i.e. 2,993 and 4,858) in MSS derived fibroblasts and LCs, respectively compared to their healthy controls. Notably, the share of altered proteins was little i.e. only seven candidates were found commonly regulated in both cell types (Fig. 5.3).



			MSS/Healthy	
UniProt accession	**Protein**	**Gene**	**Fibroblasts**	**LCs**
P21266	Glutathione S-transferase Mu 3	*GSTM3*	0.29	0.41
Q14699	Raftlin	*RFTN1*	0.45	1.40
Q9H173	Nucleotide exchange factor SIL1	*SIL1*	0.46	0.26
O43175	D-3-phosphoglycerate dehydrogenase	*PHGDH*	0.49	0.48
O60711	Leupaxin	*LPXN*	2.10	1.33
P13674	Prolyl 4-hydroxylase subunit alpha-1	*P4HA1*	2.15	0.75
P04114	Apolipoprotein B-100	*APOB*	2.34	1.65

Figure 5.3: On left, Venn-diagram depicting the overlap of differentially regulated proteins obtained from MSS fibroblasts and LCs iTRAQ datasets. Five out of seven commonly quantified proteins showed regulation in similar direction (on right). The Venn-diagram was created with http://bioinformatics.psb.ugent.be/webtools/Venn/.

Among these, five proteins including SIL1 showed regulation in the same direction. Besides SIL1, in both cell types, D-3-phosphoglycerate dehydrogenase (*PHGDH*) was also downregulated. *PHGDH* has been linked to cerebellar ataxia, cataracts and mild psychomotor retardation [193],

which are the characteristic phenotypes of MSS (see Figure 1.3). The reasons for such a minimal overlap of proteins (i.e. seven) between the two cell types can (i) be partly explained by the assumption that distinct genes are expressed in different cell populations and (ii) due to the intra-/interpatient biological variability between different MSS cases i.e. MSS2 (fibroblasts), MSS24, MSS32, MSS33, and MSS94 (all LCs). However, the abundances of BiP and GRP170 in both datasets were not significantly higher (see Sections 4.2.1 and 4.2.2) as expected since increased levels of the latter has been suggested as one of the compensatory mechanisms for the absence of SIL1 in man and mouse. One plausible explanation is that GRP170 and SIL1 might have redundant functions in certain cell types and thus GRP170 may act as an alternative NEF of BiP in the protein folding process [47, 194]. Recently, Ichhaporia et al., [195] reported that the role of SIL1/Sil1 as unnecessary in BiP-mediated antibody (immunoglobulin, Ig) production and their secretion in MSS patients (EBV-LCs) and *woozy* mice. Further, they even pointed out that overexpression of GRP170 was not seen in the SIL1-deficient mutants and the levels of antibodies remained unchanged when compared to their respective healthy controls. Surprisingly, in the MSS-LCs proteomics dataset, a negligible number of Ig-related proteins (i.e. only seven) were quantified and their relative abundances did not pass the criteria for differentially regulated proteins. The exact reason for such low identifications of immunoglobulins in the iTRAQ dataset is unclear, but the relatively stable level of GRP170 (MSS/Healthy ratio of 0.8) is in line with the findings of Ichhaporia et al.

Nevertheless, the loss of SIL1 indeed showed an effect on these apparently MSS non-vulnerable cell populations and this was observed in the proteomics data as well as in the morphological findings (Fig. 4.7 C-E), suggesting that functional SIL1 is important for maintaining overall cellular integrity. In case of MSS-fibroblasts study, upregulation of *HMOX1*, *HSPB6* and *UCHL1* indicate an activation of the UPR pathway induced by the ER stress most likely caused due to the aggregation of un-/misfolded proteins within the ER lumen. Furthermore, altered levels of proteins that belong to the major subcellular compartments including the ER, the mitochondria and the nucleus were detected in MSS-LCs. The ER and mitochondria form physical interactions involved in the regulation of mitochondrial energetics and apoptotic signaling cascades and it is becoming increasingly clear that the ER stress can also modulate mitochondrial protein homeostasis (or proteostasis) and function [196].

In MSS-LCs dataset, increased abundances of *CYCS* and *AIFM2* - both mitochondrial apoptosis promoting proteins and decreased levels of *CPT1A*, *MTRF1L*, *TOM1L2* and *TXNRD2* indicate a mitochondrial vulnerability due to SIL1 loss, which was already described in phenotypically affected tissues thus demonstrating an important interdependence of cellular SIL1 level and mitochondrial function. Besides, disturbances in the cytoskeletal organization were also detected in the MSS-LCs and moreover, initiation of the UPR and/or apoptotic pathways (for the establishment of homeostasis/cell death) was evident by the upregulation of *ANKRD13A*, *UEVLD* and *CTSB*. However, upregulation of the pro-survival proteins, such as *ACADSB*, *GPX4*, *PRDX5*, *ATXN10* and *LGALS1* could be responsible for the survival of MSS-LCs, which can be attributed to the impaired SIL1-BiP protein folding process.

5.4 *Woozy* mouse - comparable to MSS patients

Slightly over a decade ago, Zhao and colleagues [34c] described the animal model of MSS i.e. *woozy* mouse and since then it has been regarded as the broad phenocopy of human MSS [35]. Recently, Roos and co-workers [35] performed longitudinal studies on skeletal muscles derived from *woozy* to investigate the pathophysiology of Sil1 loss. Their findings showed prominent myopathic changes that progressed from mild to severe alterations in an age-dependent manner indicating the critical role of Sil1 in maintaining the skeletal muscle integrity.

On the proteome level, disturbances in different subcellular compartments most likely caused by the ER-stress were observed in both affected tissues i.e. cerebellum and skeletal muscles (see Sections 4.3.1 and 4.3.2). Remarkably, altered levels of several Ca^{2+}-associated proteins that are involved in maintaining cellular Ca^{2+} homeostasis, and proteins involved in the cytoskeletal organization were detected suggesting the impact of Sil1-loss and subsequent accumulation of un-/misfolded proteins leading to tissue damage.

In *woozy* cerebella, degeneration of the Purkinje cells accompanied by cerebral ataxia has been related with early onset of neurodegeneration in the mutant animals [34c]. Low abundances of *Calb1*, *Prkcg* and *Grid2ip*, which are involved in the development and proper functioning of this neuronal cell population, suggest that the Purkinje cells seem to be particularly sensitive to fluxes in intracellular calcium levels that could result from reduction of chaperone activity and the ER stress [197]. Nevertheless, upregulation of pro-survival protein i.e. *Kit* and proteins that are

involved in antagonizing oxidative damage including *Naprt* and *Pdrxn6* indicate that the other cerebellar cell populations might counteract the prolonged ER stress and its associated problems, including loss of calcium homeostasis. In case of *woozy* skeletal muscle, the cellular perturbations due to the absence of Sil1 were more prominent, which was evident by the upregulation of nearly 83% of the differentially altered proteins. Activation of the UPR pathway, most likely as a consequence of Sil1 loss was seen by the increased abundances *Hspa5*, *Canx* and *Pdia6*. Notably, upregulation of *Lgals3*, *Eef1a1* and *Manf* - proteins that are involved in the pro-survival and protective mechanisms, particularly in the ER-stress conditions suggest an activation of the rescue processes even in Sil1-vulnerable tissues of the *woozy*.

5.5 Preliminary insights into index patient muscle pathophysiology

Proteomics analysis of the muscle specimen derived from an index patient who manifested neurodegenerative and myopathic abnormalities revealed increased levels of several cytoskeletal proteins. There is a good amount of evidence that muscle weakness can be a consequence of gene mutations on certain cytoskeletal proteins that cause myofibrillar and myosin storage myopathies that disrupt the function of skeletal and cardiac muscles [198]. In most cases, various muscle proteins form aggregates that prevent these proteins from functioning normally in the myofibrils of the affected individuals. Therefore, based on the proteomics data it can be assumed that the index patient might have suffered from an idiopathic myopathy.

5.6 *In vitro* cell lines - suitable to study neuromuscular disorders

In vitro models (immortalized and non-immortalized) are extremely helpful to study human diseases because they allow to analyze different cell types independently from each other and to perform dynamic studies e.g. test treatments on isolated cells. Here, the human immortalized embryonic kidney cell cultures i.e. HEK293 were selected for depleting SIL1 and to subsequently investigate more closely the overall cellular disturbances upon SIL1-loss. Further, the HEK293 cells have been reported to exhibit a wide variety of differentiations including: neural crest cells, neurons and glia (of myelin sheath). In fact, HEK293 cells (i) show characteristics of immature neurons with expression of neuron-specific proteins [199], (ii) were classified as a highly efficient screening tool for the drug development in neurological diseases [200] and to study

neurodegenerative disorders [201]. Therefore, HEK293 cells were deemed fit for the generation of an *in vitro* model for MSS. The proteomics data indeed showed marked cellular perturbations in various subcellular organelles and their functions caused due to the absence of SIL1 (Fig. 4.12). These observed changes were in accordance with the findings detected in the proteomic profiling of MSS-fibroblasts and LCs. Therefore, SIL1-depleted HEK293 cells provided a better understanding of the cellular role of SIL1 in ER-associated protein processing and in manifestation and progression of neurodegenerative disorders e.g. MSS [168]. Furthermore, using the same cell line (i.e. HEK293-ΔSIL1), a PRM-based targeted MS method was established for a subset of proteins that are involved in the UPR pathway. In general, this method can be used not only to assay, but also to relatively quantify the UPR-associated proteins in different biological (human) samples wherein the activation of this pathway needs to be monitored. Notably, the assay can be performed using a single shot LC-MS/MS analysis on a MS with SRM or PRM capability e.g. Q Exactive MS.

Next, a detailed proteome characterization of the human immortal RCMH myoblastic cell line was performed to evaluate its suitability as an *in vitro* model for studying myopathies and neuromuscular related disorders since most research groups currently utilize an extensively characterized mouse myoblastic C2C12 cells for their investigations [173]. Proteome profiling of RCMH cells confirm the known expression of proteins important for muscle function. Moreover, comparison of the estimated protein abundances with two different proteomes (i.e. HeLa and U2OS), revealed that proteins related to special properties of muscle function are considerably enriched in RCMH (Table 4.3).

To conclude, this work demonstrates a detailed proteomic investigation of Marinesco-Sjögren Syndrome - a debilitating multi-system disorder caused by the impaired protein folding process. These findings highlight that the loss of functional SIL1 - even in clinically non-affected tissues of MSS patients, results in morphological perturbations that can be associated with altered protein folding machinery. Further, proteomics data revealed abnormal cytoskeletal and mitochondrial integrity, activation of antagonistic apoptotic and pro-survival mechanisms as well as altered expression of proteins necessary for function and maintenance of skeletal muscle fibers and neurons. Moreover, the quantitative proteomics data of the Sil1-deficient *woozy* mice tissues has (i) improved the knowledge about proteins that are involved in the altered pathways of the

damaged/degenerated cells, (ii) enhanced the commonality of phenotypic manifestations between MSS and *woozy* mouse due to the absence of SIL1/Sil1 and (iii) also led to the identification of some interesting candidates, which might act as pro-survival factors due to loss of functional SIL1 that have not been previously reported. Lastly, proteomic profiling of the human *in vitro* cell lines i.e. (i) HEK293-ΔSIL1 has provided deeper insights into the role of SIL1 in MSS pathophysiology and (ii) enabled to establish RCMH as a suitable model to investigate processes related to muscle function e.g. mechanical stress burden and mechanotransduction, EC-coupling and other ER-associated myopathic disorders.

6 References

1. Andersen, G. R.; Nissen, P.; Nyborg, J., Elongation factors in protein biosynthesis. *Trends in Biochemical Sciences* **2003,** *28* (8), 434-441.
2. Kleizen, B.; Braakman, I., Protein folding and quality control in the endoplasmic reticulum. *Current Opinion in Cell Biology* **2004,** *16* (4), 343-349.
3. Hebert, D. N.; Molinari, M., In and Out of the ER: Protein Folding, Quality Control, Degradation, and Related Human Diseases. *Physiological Reviews* **2007,** *87* (4), 1377-1408.
4. Vitale, A.; Ceriotti, A.; Denecke, J., The Role of the Endoplasmic Reticulum in Protein Synthesis, Modification and Intracellular Transport. *Journal of Experimental Botany* **1993,** *44* (9), 1417-1444.
5. Zimmermann, R.; Mueller, L.; Wullich, B., Protein transport into the endoplasmic reticulum: mechanisms and pathologies. *Trends in Molecular Medicine* **2006,** *12* (12), 567-573.
6. Johnson, A. E.; van Waes, M. A., THE TRANSLOCON: A Dynamic Gateway at the ER Membrane. *Annual Review of Cell and Developmental Biology* **1999,** *15* (1), 799-842.
7. Hartl, F. U., Molecular chaperones in cellular protein folding. *Nature* **1996,** *381* (6583), 571-580.
8. Meldolesi, J.; Pozzan, T., The endoplasmic reticulum Ca2+ store: a view from the lumen. *Trends in Biochemical Sciences* **1998,** *23* (1), 10-14.
9. Bygrave, F. L.; Benedetti, A., What is the concentration of calcium ions in the endoplasmic reticulum? *Cell Calcium* **1996,** *19* (6), 547-551.
10. Behnke, J.; Feige, M. J.; Hendershot, L. M., BiP and Its Nucleotide Exchange Factors Grp170 and Sil1: Mechanisms of Action and Biological Functions. *Journal of Molecular Biology* **2015,** *427* (7), 1589-1608.
11. (a) Ellgaard, L.; Helenius, A., Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol* **2003,** *4* (3), 181-191; (b) Schröder, M.; Kaufman, R. J., ER stress and the unfolded protein response. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **2005,** *569* (1–2), 29-63.
12. (a) Haas, I. G.; Wabl, M., Immunoglobulin heavy chain binding protein. *Nature* **1983,** *306* (5941), 387-389; (b) Munro, S.; Pelham, H. R. B., An hsp70-like protein in the ER: Identity with the 78 kd glucose-regulated protein and immunoglobulin heavy chain binding protein. *Cell* **1986,** *46* (2), 291-300.
13. Hendershot, L. M., The ER function BiP is a master regulator of ER function. *Mt Sinai J Med* **2004,** *71* (5), 289-97.
14. Mayer, M. P.; Bukau, B., Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences* **2005,** *62* (6), 670-684.
15. (a) Tyson, J. R.; Stirling, C. J., LHS1 and SIL1 provide a lumenal function that is essential for protein translocation into the endoplasmic reticulum. *EMBO J* **2000,** *19* (23), 6440-52; (b) Hale, S. J.; Lovell, S. C.; de Keyzer, J.; Stirling, C. J., Interactions between Kar2p and Its Nucleotide Exchange Factors Sil1p and Lhs1p Are Mechanistically Distinct. *Journal of Biological Chemistry* **2010,** *285* (28), 21600-21606; (c) Chung, K. T.; Shen, Y.; Hendershot, L. M., BAP, a mammalian BiP-associated protein, is a nucleotide exchange factor that regulates the ATPase activity of BiP. *J Biol Chem* **2002,** *277* (49), 47557-63; (d) Steel, G. J.; Fullerton, D. M.; Tyson, J. R.; Stirling, C. J., Coordinated Activation of Hsp70 Chaperones. *Science* **2004,** *303* (5654), 98-101.

References

16. Zahedi, R. P.; Völzing, C.; Schmitt, A.; Frien, M.; Jung, M.; Dudek, J.; Wortelkamp, S.; Sickmann, A.; Zimmermann, R., Analysis of the membrane proteome of canine pancreatic rough microsomes identifies a novel Hsp40, termed ERj7. *PROTEOMICS* **2009,** *9* (13), 3463-3473.
17. Andréasson, C.; Rampelt, H.; Fiaux, J.; Druffel-Augustin, S.; Bukau, B., The Endoplasmic Reticulum Grp170 Acts as a Nucleotide Exchange Factor of Hsp70 via a Mechanism Similar to That of the Cytosolic Hsp110. *Journal of Biological Chemistry* **2010,** *285* (16), 12445-12453.
18. Zoghbi, H. Y., SILencing misbehaving proteins. *Nat Genet* **2005,** *37* (12), 1302-3.
19. Yan, M.; Li, J.; Sha, B., Structural analysis of the Sil1-Bip complex reveals the mechanism for Sil1 to function as a nucleotide-exchange factor. *Biochemical Journal* **2011,** *438*, 447-455.
20. Melnyk, A.; Rieger, H.; Zimmermann, R., Co-chaperones of the mammalian endoplasmic reticulum. In *The Networking of Chaperones by Co-chaperones*, Springer: 2015; pp 179-200.
21. Rutkowski, D. T.; Kaufman, R. J., A trip to the ER: coping with stress. *Trends in Cell Biology* **2004,** *14* (1), 20-28.
22. (a) Ni, M.; Lee, A. S., ER chaperones in mammalian development and human diseases. *FEBS letters* **2007,** *581* (19), 3641-3651; (b) Schröder, M.; Kaufman, R. J., THE MAMMALIAN UNFOLDED PROTEIN RESPONSE. *Annual Review of Biochemistry* **2005,** *74* (1), 739-789.
23. Kopito, R. R., Aggresomes, inclusion bodies and protein aggregation. *Trends in Cell Biology* **2000,** *10* (12), 524-530.
24. Fink, A. L., Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding and Design* **1998,** *3* (1), R9-R23.
25. Walter, P.; Ron, D., The Unfolded Protein Response: From Stress Pathway to Homeostatic Regulation. *Science* **2011,** *334* (6059), 1081-1086.
26. Kim, I.; Xu, W.; Reed, J. C., Cell death and endoplasmic reticulum stress: disease relevance and therapeutic opportunities. *Nat Rev Drug Discov* **2008,** *7* (12), 1013-1030.
27. Ron, D.; Walter, P., Signal integration in the endoplasmic reticulum unfolded protein response. *Nat Rev Mol Cell Biol* **2007,** *8* (7), 519-529.
28. Wang, M.; Ye, R.; Barron, E.; Baumeister, P.; Mao, C.; Luo, S.; Fu, Y.; Luo, B.; Dubeau, L.; Hinton, D. R.; Lee, A. S., Essential role of the unfolded protein response regulator GRP78/BiP in protection from neuronal apoptosis. *Cell Death and Differentiation* **2010,** *17* (3), 488-498.
29. Shen, J.; Chen, X.; Hendershot, L.; Prywes, R., ER Stress Regulation of ATF6 Localization by Dissociation of BiP/GRP78 Binding and Unmasking of Golgi Localization Signals. *Developmental Cell* **2002,** *3* (1), 99-111.
30. Meusser, B.; Hirsch, C.; Jarosch, E.; Sommer, T., ERAD: the long road to destruction. *Nat Cell Biol* **2005,** *7* (8), 766-772.
31. Voges, D.; Zwickl, P.; Baumeister, W., The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu Rev Biochem* **1999,** *68*, 1015-68.
32. Ciechanover, A., The ubiquitin-proteasome pathway: on protein death and cell life. *The EMBO Journal* **1998,** *17* (24), 7151-7160.
33. Szegezdi, E.; Logue, S. E.; Gorman, A. M.; Samali, A., Mediators of endoplasmic reticulum stress-induced apoptosis. *EMBO Reports* **2006,** *7* (9), 880-885.
34. (a) Ross, C. A.; Poirier, M. A., Protein aggregation and neurodegenerative disease. *Nat Med* **2004,** *10 Suppl*, S10-7; (b) Yoshida, H., ER stress and diseases. *FEBS Journal* **2007,** *274* (3), 630-658; (c) Zhao, L.; Longo-Guess, C.; Harris, B. S.; Lee, J. W.; Ackerman, S. L., Protein accumulation and neurodegeneration in the woozy mutant mouse is caused by disruption of SIL1, a cochaperone of BiP. *Nat Genet* **2005,** *37*, 974-979.

35. Roos, A.; Buchkremer, S.; Kollipara, L.; Labisch, T.; Gatz, C.; Zitzelsberger, M.; Brauers, E.; Nolte, K.; Schroder, J. M.; Kirschner, J.; Jesse, C. M.; Goebel, H. H.; Goswami, A.; Zimmermann, R.; Zahedi, R. P.; Senderek, J.; Weis, J., Myopathy in Marinesco-Sjogren syndrome links endoplasmic reticulum chaperone dysfunction to nuclear envelope pathology. *Acta neuropathologica* **2014,** *127* (5), 761-77.
36. Carrell, R. W.; Lomas, D. A., Conformational disease. *The Lancet* **1997,** *350* (9071), 134-138.
37. Ni, M.; Lee, A. S., ER chaperones in mammalian development and human diseases. *FEBS Lett* **2007,** *581*, 3641-3651.
38. Doyle, K. M.; Kennedy, D.; Gorman, A. M.; Gupta, S.; Healy, S. J.; Samali, A., Unfolded proteins and endoplasmic reticulum stress in neurodegenerative disorders. *Journal of cellular and molecular medicine* **2011,** *15* (10), 2025-39.
39. Marinesco, G., Draganesco, S., Vasiliu, D., Nouvelle maladie familiale caracterisee par une cataracte congenitale et un arret du development somato-neuro-psychique. *Encephale 26* **1931**, 97-109.
40. Anttonen, A.-K., The molecular basis of Marinesco-Sjögren syndrome. **2008**.
41. (a) Krieger, M.; Roos, A.; Stendel, C.; Claeys, K. G.; Sonmez, F. M.; Baudis, M.; Bauer, P.; Bornemann, A.; de Goede, C.; Dufke, A.; Finkel, R. S.; Goebel, H. H.; Haussler, M.; Kingston, H.; Kirschner, J.; Medne, L.; Muschke, P.; Rivier, F.; Rudnik-Schoneborn, S.; Spengler, S.; Inzana, F.; Stanzial, F.; Benedicenti, F.; Synofzik, M.; Lia Taratuto, A.; Pirra, L.; Tay, S. K.; Topaloglu, H.; Uyanik, G.; Wand, D.; Williams, D.; Zerres, K.; Weis, J.; Senderek, J., SIL1 mutations and clinical spectrum in patients with Marinesco-Sjogren syndrome. *Brain* **2013,** *136* (Pt 12), 3634-44; (b) Senderek, J.; Krieger, M.; Stendel, C.; Bergmann, C.; Moser, M.; Breitbach-Faller, N.; Rudnik-Schoneborn, S.; Blaschek, A.; Wolf, N. I.; Harting, I.; North, K.; Smith, J.; Muntoni, F.; Brockington, M.; Quijano-Roy, S.; Renault, F.; Herrmann, R.; Hendershot, L. M.; Schroder, J. M.; Lochmuller, H.; Topaloglu, H.; Voit, T.; Weis, J.; Ebinger, F.; Zerres, K., Mutations in SIL1 cause Marinesco-Sjogren syndrome, a cerebellar ataxia with cataract and myopathy. *Nature Genetics* **2005,** *37* (12), 1312-1314; (c) Anttonen, A.-K.; Siintola, E.; Tranebjaerg, L.; Iwata, N. K.; Bijlsma, E. K.; Meguro, H.; Ichikawa, Y.; Goto, J.; Kopra, O.; Lehesjoki, A.-E., Novel SIL1 mutations and exclusion of functional candidate genes in Marinesco-Sjogren syndrome. *Eur J Hum Genet* **2008,** *16* (8), 961-969.
42. Anttonen, A. K.; Mahjneh, I.; Hamalainen, R. H.; Lagier-Tourenne, C.; Kopra, O.; Waris, L.; Anttonen, M.; Joensuu, T.; Kalimo, H.; Paetau, A.; Tranebjaerg, L.; Chaigne, D.; Koenig, M.; Eeg-Olofsson, O.; Udd, B.; Somer, M.; Somer, H.; Lehesjoki, A. E., The gene disrupted in Marinesco-Sjogren syndrome encodes SIL1, an HSPA5 cochaperone. *Nat Genet* **2005,** *37* (12), 1309-11.
43. Sewry, C. A.; Voit, T.; Dubowitz, V., Myopathy with unique ultrastructural feature in Marinesco-Sjogren syndrome. *Annals of neurology* **1988,** *24* (4), 576-80.
44. Cerami, C.; Tarantino, P.; Cupidi, C.; Annesi, G.; Lo Re, V.; Gagliardi, M.; Piccoli, T.; Quattrone, A., Marinesco–Sjögren syndrome caused by a new SIL1 frameshift mutation. *Journal of the Neurological Sciences* **2015,** *354* (1–2), 112-113.
45. (a) Gidalevitz, T.; Stevens, F.; Argon, Y., Orchestration of secretory protein folding by ER chaperones. *Biochimica Et Biophysica Acta-Molecular Cell Research* **2013,** *1833* (11), 2410-2424; (b) Filezac de L'Etang, A.; Maharjan, N.; Cordeiro Brana, M.; Ruegsegger, C.; Rehmann, R.; Goswami, A.; Roos, A.; Troost, D.; Schneider, B. L.; Weis, J.; Saxena, S., Marinesco-Sjogren

syndrome protein SIL1 regulates motor neuron subtype-selective ER stress in ALS. *Nat Neurosci* **2015,** *18* (2), 227-238.
46. Buchkremer, S. G., C; José, Andrésa; Weis, J; Roos, Andreas, Sil1-Mutant Mice Elucidate Chaperone Function in Neurological Disorders. *Journal of Neuromuscular Diseases* **2016,** *3* (2), 169-181.
47. Zhao, L.; Rosales, C.; Seburn, K.; Ron, D.; Ackerman, S. L., Alteration of the unfolded protein response modifies neurodegeneration in a mouse model of Marinesco-Sjogren syndrome. *Human Molecular Genetics* **2010,** *19* (1), 25-35.
48. Inaguma, Y.; Hamada, N.; Tabata, H.; Iwamoto, I.; Mizuno, M.; Nishimura, Y. V.; Ito, H.; Morishita, R.; Suzuki, M.; Ohno, K.; Kumagai, T.; Nagata, K., SIL1, a causative cochaperone gene of Marinesco-Sojgren syndrome, plays an essential role in establishing the architecture of the developing cerebral cortex. *EMBO molecular medicine* **2014,** *6* (3), 414-29.
49. Rappsilber, J.; Mann, M., What does it mean to identify a protein in proteomics? *Trends in Biochemical Sciences* **2002,** *27* (2), 74-78.
50. Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I.; Williams, K. L.; Hochstrasser, D. F., From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio/technology (Nature Publishing Company)* **1996,** *14* (1), 61-5.
51. Human Genome Sequencing, C., Finishing the euchromatic sequence of the human genome. *Nature* **2004,** *431* (7011), 931-945.
52. Pertea, M.; Salzberg, S. L., Between a chicken and a grape: estimating the number of human genes. *Genome Biology* **2010,** *11* (5), 206-206.
53. Mirza, S. P.; Olivier, M., Methods and approaches for the comprehensive characterization and quantification of cellular proteomes using mass spectrometry. *Physiological genomics* **2008,** *33* (1), 3-11.
54. Legrain, P.; Aebersold, R.; Archakov, A.; Bairoch, A.; Bala, K.; Beretta, L.; Bergeron, J.; Borchers, C. H.; Corthals, G. L.; Costello, C. E.; Deutsch, E. W.; Domon, B.; Hancock, W.; He, F.; Hochstrasser, D.; Marko-Varga, G.; Salekdeh, G. H.; Sechi, S.; Snyder, M.; Srivastava, S.; Uhlén, M.; Wu, C. H.; Yamamoto, T.; Paik, Y.-K.; Omenn, G. S., The Human Proteome Project: Current State and Future Direction. *Molecular & Cellular Proteomics* **2011,** *10* (7).
55. Nagaraj, N.; Wisniewski, J. R.; Geiger, T.; Cox, J.; Kircher, M.; Kelso, J.; Pääbo, S.; Mann, M., Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular Systems Biology* **2011,** *7* (1).
56. Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J., The One Hour Yeast Proteome. *Molecular & Cellular Proteomics* **2014,** *13* (1), 339-347.
57. (a) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry* **2007,** *389* (4), 1017-1031; (b) Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics* **2002,** *1* (5), 376-386; (c) Wilm, M., Quantitative proteomics in biological research. *PROTEOMICS* **2009,** *9* (20), 4590-4605.
58. Hanash, S., Disease proteomics. *Nature* **2003,** *422* (6928), 226-232.

59. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry* **1988,** *60* (20), 2299-2301.
60. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** *246* (4926), 64-71.
61. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** *422* (6928), 198-207.
62. Wilm, M. S.; Mann, M., Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *International Journal of Mass Spectrometry and Ion Processes* **1994,** *136* (2), 167-180.
63. (a) Wilm, M.; Mann, M., Analytical properties of the nanoelectrospray ion source. *Anal Chem* **1996,** *68* (1), 1-8; (b) Karas, M.; Bahr, U.; Dülcks, T., Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. *Fresenius J Anal Chem* **2000,** *366* (6-7), 669-676.
64. Zhang, X.; Fang, A.; Riley, C. P.; Wang, M.; Regnier, F. E.; Buck, C., Multi-dimensional Liquid Chromatography in Proteomics. *Analytica chimica acta* **2010,** *664* (2), 101-113.
65. Michel, P. E.; Reymond, F.; Arnaud, I. L.; Josserand, J.; Girault, H. H.; Rossier, J. S., Protein fractionation in a multicompartment device using Off-Gel™ isoelectric focusing. *ELECTROPHORESIS* **2003,** *24* (1-2), 3-11.
66. Delahunty, C.; Yates Iii, J. R., Protein identification using 2D-LC-MS/MS. *Methods* **2005,** *35* (3), 248-255.
67. (a) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd, Direct analysis of protein complexes using mass spectrometry. *Nature biotechnology* **1999,** *17* (7), 676-82; (b) Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotech* **2001,** *19* (3), 242-247.
68. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C., Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *Journal of Separation Science* **2005,** *28* (14), 1694-1703.
69. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C., Orthogonality of Separation in Two-Dimensional Liquid Chromatography. *Analytical Chemistry* **2005,** *77* (19), 6426-6434.
70. (a) Yang, F.; Shen, Y.; Camp, D. G.; Smith, R. D., High pH reversed-phase chromatography with fraction concatenation as an alternative to strong-cation exchange chromatography for two-dimensional proteomic analysis. *Expert Review of Proteomics* **2012,** *9* (2), 129-134; (b) Wang, Y.; Yang, F.; Gritsenko, M. A.; Wang, Y.; Clauss, T.; Liu, T.; Shen, Y.; Monroe, M. E.; Lopez-Ferrer, D.; Reno, T.; Moore, R. J.; Klemke, R. L.; Camp, D. G.; Smith, R. D., Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* **2011,** *11* (10), 2019-2026.
71. Mann, M.; Kelleher, N. L., Precision proteomics: The case for high resolution and high mass accuracy. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** *105* (47), 18132-18138.
72. Makarov, A., Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry* **2000,** *72* (6), 1156-1162.
73. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R., The Orbitrap: a new mass spectrometer. *Journal of mass spectrometry : JMS* **2005,** *40* (4), 430-43.

74. Schwartz, J. C.; Senko, M. W.; Syka, J. E. P., A two-dimensional quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry* **2002,** *13* (6), 659-669.
75. (a) March, R. E., An Introduction to Quadrupole Ion Trap Mass Spectrometry. *Journal of Mass Spectrometry* **1997,** *32* (4), 351-369; (b) Miller, P. E.; Denton, M. B., The quadrupole mass filter: Basic operating concepts. *Journal of Chemical Education* **1986,** *63* (7), 617.
76. Kingdon, K. H., A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Physical Review* **1923,** *21* (4), 408-418.
77. Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Müller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M.; Dunyach, J.-J.; Cox, J.; Horning, S.; Mann, M.; Makarov, A., Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Molecular & Cellular Proteomics : MCP* **2012,** *11* (3), O111.013698.
78. Michalski, A.; Damoc, E.; Hauschild, J.-P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S., Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics* **2011,** *10* (9).
79. Steen, H.; Mann, M., The abc's (and xyz's) of peptide sequencing. *Nat Rev Mol Cell Biol* **2004,** *5* (9), 699-711.
80. Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R., Protein sequencing by tandem mass spectrometry. *Proceedings of the National Academy of Sciences* **1986,** *83* (17), 6233-6237.
81. McLuckey, S. A., Principles of collisional activation in analytical mass spectrometry. *Journal of the American Society for Mass Spectrometry* **1992,** *3* (6), 599-614.
82. Roepstorff, P.; Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry* **1984,** *11* (11), 601.
83. Papayannopoulos, I. A., The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrometry Reviews* **1995,** *14* (1), 49-73.
84. Johnson, J. V.; Yost, R. A.; Kelley, P. E.; Bradford, D. C., Tandem-in-space and tandem-in-time mass spectrometry: triple quadrupoles and quadrupole ion traps. *Analytical Chemistry* **1990,** *62* (20), 2162-2172.
85. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nat Meth* **2007,** *4* (9), 709-712.
86. McAlister, G. C.; Phanstiel, D. H.; Brumbaugh, J.; Westphall, M. S.; Coon, J. J., Higher-energy Collision-activated Dissociation Without a Dedicated Collision Cell. *Molecular & Cellular Proteomics : MCP* **2011,** *10* (5), O111.009456.
87. Fonslow, B. R.; Carvalho, P. C.; Academia, K.; Freeby, S.; Xu, T.; Nakorchevsky, A.; Paulus, A.; Yates, J. R., Improvements in proteomic metrics of low abundance proteins through proteome equalization using ProteoMiner prior to MudPIT. *Journal of proteome research* **2011,** *10* (8), 3690-3700.
88. Venable, J. D.; Dong, M.-Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R., Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Meth* **2004,** *1* (1), 39-45.
89. Eng, J. K.; McCormack, A. L.; Yates, J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **1994,** *5* (11), 976-989.

90. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** *20* (18), 3551-67.
91. Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Meth* **2007,** *4* (3), 207-214.
92. Nesvizhskii, A. I., A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of Proteomics* **2010,** *73* (11), 2092-2123.
93. Aebersold, R., Quantitative Proteome Analysis: Methods and Applications. *Journal of Infectious Diseases* **2003,** *187* (Supplement 2), S315-S320.
94. (a) Hawkridge, A. M., CHAPTER 1 Practical Considerations and Current Limitations in Quantitative Mass Spectrometry-based Proteomics. In *Quantitative Proteomics*, The Royal Society of Chemistry: 2014; pp 1-25; (b) Nikolov, M.; Schmidt, C.; Urlaub, H., Quantitative mass spectrometry-based proteomics: an overview. *Methods in molecular biology (Clifton, N.J.)* **2012,** *893*, 85-100.
95. Becker, G. W., Stable isotopic labeling of proteins for quantitative proteomic applications. *Briefings in Functional Genomics & Proteomics* **2008,** *7* (5), 371-382.
96. Gevaert, K.; Impens, F.; Ghesquiere, B.; Van Damme, P.; Lambrechts, A.; Vandekerckhove, J., Stable isotopic labeling in proteomics. *Proteomics* **2008,** *8* (23-24), 4873-85.
97. Gouw, J. W.; Krijgsveld, J.; Heck, A. J. R., Quantitative Proteomics by Metabolic Labeling of Model Organisms. *Molecular & Cellular Proteomics : MCP* **2010,** *9* (1), 11-24.
98. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Molecular & Cellular Proteomics* **2004,** *3* (12), 1154-1169.
99. Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C., Proteolytic 18O Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus. *Analytical Chemistry* **2001,** *73* (13), 2836-2842.
100. Liebler, D. C.; Zimmerman, L. J., Targeted Quantitation of Proteins by Mass Spectrometry. *Biochemistry* **2013,** *52* (22), 3797-3806.
101. Unwin, R., Quantification of Proteins by iTRAQ. In *LC-MS/MS in Proteomics*, Cutillas, P. R.; Timms, J. F., Eds. Humana Press: 2010; Vol. 658, pp 205-215.
102. Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Wright, P. C., Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation. *PROTEOMICS* **2011,** *11* (11), 2341-2346.
103. (a) Mahoney, D. W.; Therneau, T. M.; Heppelmann, C. J.; Higgins, L.; Benson, L. M.; Zenka, R. M.; Jagtap, P.; Nelsestuen, G. L.; Bergen, H. R.; Oberg, A. L., Relative Quantification: Characterization of Bias, Variability and Fold Changes in Mass Spectrometry Data from iTRAQ-Labeled Peptides. *Journal of Proteome Research* **2011,** *10* (9), 4325-4333; (b) Christoforou, A.; Lilley, K., Isobaric tagging approaches in quantitative proteomics: the ups and downs. *Analytical and Bioanalytical Chemistry* **2012,** *404* (4), 1029-1037.
104. Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Rehman, I.; Wright, P. C., iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J Proteome Res* **2009,** *8* (11), 5347-55.

References

105. Savitski, M. M.; Mathieson, T.; Zinn, N.; Sweetman, G.; Doce, C.; Becher, I.; Pachl, F.; Kuster, B.; Bantscheff, M., Measuring and Managing Ratio Compression for Accurate iTRAQ/TMT Quantification. *Journal of Proteome Research* **2013,** *12* (8), 3586-3598.
106. Ting, L.; Rad, R.; Gygi, S. P.; Haas, W., MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Meth* **2011,** *8* (11), 937-940.
107. Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G., Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics. *Molecular & Cellular Proteomics* **2005,** *4* (10), 1487-1502.
108. Neilson, K. A.; Ali, N. A.; Muralidharan, S.; Mirzaei, M.; Mariani, M.; Assadourian, G.; Lee, A.; van Sluyter, S. C.; Haynes, P. A., Less label, more free: Approaches in label-free quantitative mass spectrometry. *PROTEOMICS* **2011,** *11* (4), 535-553.
109. Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M., Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Molecular & Cellular Proteomics* **2005,** *4* (9), 1265-1272.
110. Braisted, J. C.; Kuntumalla, S.; Vogel, C.; Marcotte, E. M.; Rodrigues, A. R.; Wang, R.; Huang, S.-T.; Ferlanti, E. S.; Saeed, A. I.; Fleischmann, R. D.; Peterson, S. N.; Pieper, R., The APEX Quantitative Proteomics Tool: Generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics* **2008,** *9*, 529-529.
111. (a) Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P., Statistical Analysis of Membrane Proteome Expression Changes in Saccharomyces cerevisiae. *Journal of Proteome Research* **2006,** *5* (9), 2339-2347; (b) Paoletti, A. C.; Parmely, T. J.; Tomomori-Sato, C.; Sato, S.; Zhu, D.; Conaway, R. C.; Conaway, J. W.; Florens, L.; Washburn, M. P., Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** *103* (50), 18928-18933.
112. McIlwain, S.; Mathews, M.; Bereman, M. S.; Rubel, E. W.; MacCoss, M. J.; Noble, W. S., Estimating relative abundances of proteins from shotgun proteomics data. *BMC Bioinformatics* **2012,** *13*, 308-308.
113. Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J., Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Molecular & Cellular Proteomics* **2012,** *11* (11), 1475-1488.
114. Ronsein, G. E.; Pamir, N.; von Haller, P. D.; Kim, D. S.; Oda, M. N.; Jarvik, G. P.; Vaisar, T.; Heinecke, J. W., Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *Journal of Proteomics* **2015,** *113*, 388-399.
115. Lange, V.; Picotti, P.; Domon, B.; Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology* **2008,** *4*, 222-222.
116. Gallien, S.; Duriez, E.; Crone, C.; Kellmann, M.; Moehring, T.; Domon, B., Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol Cell Proteomics* **2012,** *11* (12), 1709-23.
117. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010,** *26* (7), 966-968.

118. Caviedes, R.; Liberona, J. L.; Hidalgo, J.; Tascon, S.; Salas, K.; Jaimovich, E., A human skeletal muscle cell line obtained from an adult donor. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1992,** *1134* (3), 247-255.
119. Manza, L. L.; Stamer, S. L.; Ham, A.-J. L.; Codreanu, S. G.; Liebler, D. C., Sample preparation and digestion for proteomic analyses using spin filters. *PROTEOMICS* **2005,** *5* (7), 1742-1745.
120. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Meth* **2009,** *6* (5), 359-362.
121. Wiśniewski, J. R.; Zielinska, D. F.; Mann, M., Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method. *Analytical Biochemistry* **2011,** *410* (2), 307-309.
122. Kollipara, L.; Zahedi, R. P., Protein carbamylation: In vivo modification or in vitro artefact? *PROTEOMICS* **2013,** *13* (6), 941-944.
123. Gilliland, G. L.; Teplyakov, A., Structural Calcium (Trypsin, Subtilisin). In *Handbook of Metalloproteins*, John Wiley & Sons, Ltd: 2006.
124. Burkhart, J. M.; Schumbrutzki, C.; Wortelkamp, S.; Sickmann, A.; Zahedi, R. P., Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *Journal of Proteomics* **2012,** *75* (4), 1454-1462.
125. Toll, H.; Oberacher, H.; Swart, R.; Huber, C. G., Separation, detection, and identification of peptides by ion-pair reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry at high and low pH. *J Chromatogr A* **2005,** *1079* (1-2), 274-86.
126. (a) Batth, T. S.; Francavilla, C.; Olsen, J. V., Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *J Proteome Res* **2014**; (b) Castillo, A.; Roig-Navarro, A. F.; Pozo, O. J., Secondary interactions, an unexpected problem emerged between hydroxyl containing analytes and fused silica capillaries in anion-exchange micro-liquid chromatography. *Journal of Chromatography A* **2007,** *1172* (2), 179-185.
127. Choksawangkarn, W.; Edwards, N.; Wang, Y.; Gutierrez, P.; Fenselau, C., Comparative Study of Workflows Optimized for In-gel, In-solution, and On-filter Proteolysis in the Analysis of Plasma Membrane Proteins. *Journal of proteome research* **2012,** *11* (5), 3030-3034.
128. Helbig, A. O.; Gauci, S.; Raijmakers, R.; van Breukelen, B.; Slijper, M.; Mohammed, S.; Heck, A. J. R., Profiling of N-Acetylated Protein Termini Provides In-depth Insights into the N-terminal Nature of the Proteome. *Molecular & Cellular Proteomics* **2010,** *9* (5), 928-939.
129. Yang, B.; Wu, Y.-J.; Zhu, M.; Fan, S.-B.; Lin, J.; Zhang, K.; Li, S.; Chi, H.; Li, Y.-X.; Chen, H.-F.; Luo, S.-K.; Ding, Y.-H.; Wang, L.-H.; Hao, Z.; Xiu, L.-Y.; Chen, S.; Ye, K.; He, S.-M.; Dong, M.-Q., Identification of cross-linked peptides from complex samples. *Nat Meth* **2012,** *advance online publication*.
130. Olsen, J. V.; de Godoy, L. M. F.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M., Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. *Molecular & Cellular Proteomics* **2005,** *4* (12), 2010-2021.
131. Thingholm, T. E.; Palmisano, G.; Kjeldsen, F.; Larsen, M. R., Undesirable Charge-Enhancement of Isobaric Tagged Phosphopeptides Leads to Reduced Identification Efficiency. *Journal of Proteome Research* **2010,** *9* (8), 4045-4052.
132. (a) Lee, A. S., Glucose-regulated proteins in cancer: molecular mechanisms and therapeutic potential. *Nat Rev Cancer* **2014,** *14* (4), 263-276; (b) Greiner, M.; Kreutzer, B.; Lang,

S.; Jung, V.; Cavalié, A.; Unteregger, G.; Zimmermann, R.; Wullich, B., Sec62 protein level is crucial for the ER stress tolerance of prostate cancer. *The Prostate* **2011,** *71* (10), 1074-1083.
133. Vaudel, M.; Barsnes, H.; Berven, F. S.; Sickmann, A.; Martens, L., SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *PROTEOMICS* **2011,** *11* (5), 996-999.
134. Vaudel, M.; Burkhart, J. M.; Zahedi, R. P.; Oveland, E.; Berven, F. S.; Sickmann, A.; Martens, L.; Barsnes, H., PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nature biotechnology* **2015,** *33* (1), 22-4.
135. Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P., ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **2008,** *24* (21), 2534-2536.
136. Muth, T.; Vaudel, M.; Barsnes, H.; Martens, L.; Sickmann, A., XTandem Parser: An open-source library to parse and analyse X!Tandem MS/MS search results. *PROTEOMICS* **2010,** *10* (7), 1522-1524.
137. (a) Kim, S.; Gupta, N.; Pevzner, P. A., Spectral Probabilities and Generating Functions of Tandem Mass Spectra: A Strike against Decoy Databases. *Journal of Proteome Research* **2008,** *7* (8), 3354-3363; (b) Kim, S.; Mischerikow, N.; Bandeira, N.; Navarro, J. D.; Wich, L.; Mohammed, S.; Heck, A. J. R.; Pevzner, P. A., The Generating Function of CID, ETD, and CID/ETD Pairs of Tandem Mass Spectra: Applications to Database Search. *Molecular & Cellular Proteomics* **2010,** *9* (12), 2840-2852.
138. Nahnsen, S.; Bielow, C.; Reinert, K.; Kohlbacher, O., Tools for Label-free Peptide Quantification(). *Molecular & Cellular Proteomics : MCP* **2013,** *12* (3), 549-556.
139. Stark, G. R.; Stein, W. H.; Moore, S., Reactions of the Cyanate Present in Aqueous Urea with Amino Acids and Proteins. *Journal of Biological Chemistry* **1960,** *235* (11), 3177-3181.
140. Stark, G. R.; Smyth, D. G., The Use of Cyanate for the Determination of NH2-terminal Residues in Proteins. *Journal of Biological Chemistry* **1963,** *238* (1), 214-226.
141. Ezgu, F.; Krejci, P.; Li, S.; de Sousa, C.; Graham, J. M., Jr.; Hansmann, I.; He, W.; Porpora, K.; Wand, D.; Wertelecki, W.; Schneider, A.; Wilcox, W. R., Phenotype-genotype correlations in patients with Marinesco-Sjogren syndrome. *Clin Genet* **2014,** *86* (1), 74-84.
142. (a) Zahedi, R. P.; Volzing, C.; Schmitt, A.; Frien, M.; Jung, M.; Dudek, J.; Wortelkamp, S.; Sickmann, A.; Zimmermann, R., Analysis of the membrane proteome of canine pancreatic rough microsomes identifies a novel Hsp40, termed ERj7. *Proteomics* **2009,** *9* (13), 3463-73; (b) Weitzmann, A.; Baldes, C.; Dudek, J.; Zimmermann, R., The heat shock protein 70 molecular chaperone network in the pancreatic endoplasmic reticulum - a quantitative approach. *The FEBS journal* **2007,** *274* (19), 5175-87.
143. Harding, H. P.; Zhang, Y.; Zeng, H.; Novoa, I.; Lu, P. D.; Calfon, M.; Sadri, N.; Yun, C.; Popko, B.; Paules, R.; Stojdl, D. F.; Bell, J. C.; Hettmann, T.; Leiden, J. M.; Ron, D., An Integrated Stress Response Regulates Amino Acid Metabolism and Resistance to Oxidative Stress. *Molecular Cell* **2003,** *11* (3), 619-633.
144. Byrne, S.; Dlamini, N.; Lumsden, D.; Pitt, M.; Zaharieva, I.; Muntoni, F.; King, A.; Robert, L.; Jungbluth, H., SIL1-related Marinesco–Sjoegren syndrome (MSS) with associated motor neuronopathy and bradykinetic movement disorder. *Neuromuscular Disorders* **2015,** *25* (7), 585-588.
145. Hashimoto, M.; Rockenstein, E.; Crews, L.; Masliah, E., Role of protein aggregation in mitochondrial dysfunction and neurodegeneration in Alzheimer's and Parkinson's diseases. *Neuromolecular medicine* **2003,** *4* (1-2), 21-36.

References

146. (a) Gurel, P. S.; Hatch, A. L.; Higgs, H. N., Connecting the cytoskeleton to the endoplasmic reticulum and Golgi. *Current biology : CB* **2014,** *24* (14), R660-72; (b) Liu, Z.; Du, X.; Deng, J.; Gu, M.; Hu, H.; Gui, M.; Yin, C. C.; Chang, Z., The interactions between mitochondria and sarcoplasmic reticulum and the proteome characterization of mitochondrion-associated membrane from rabbit skeletal muscle. *Proteomics* **2015,** *15* (15), 2701-4.
147. Lim, J.; Hao, T.; Shaw, C.; Patel, A. J.; Szabó, G.; Rual, J.-F.; Fisk, C. J.; Li, N.; Smolyar, A.; Hill, D. E.; Barabási, A.-L.; Vidal, M.; Zoghbi, H. Y., A Protein–Protein Interaction Network for Human Inherited Ataxias and Disorders of Purkinje Cell Degeneration. *Cell 125* (4), 801-814.
148. Camby, I.; Le Mercier, M.; Lefranc, F.; Kiss, R., Galectin-1: a small protein with major functions. *Glycobiology* **2006,** *16* (11), 137R-157R.
149. Starossom, Sarah C.; Mascanfroni, Ivan D.; Imitola, J.; Cao, L.; Raddassi, K.; Hernandez, Silvia F.; Bassil, R.; Croci, Diego O.; Cerliani, Juan P.; Delacour, D.; Wang, Y.; Elyaman, W.; Khoury, Samia J.; Rabinovich, Gabriel A., Galectin-1 Deactivates Classically Activated Microglia and Protects from Inflammation-Induced Neurodegeneration. *Immunity* **2012,** *37* (2), 249-263.
150. Lemon, R. N.; Edgley, S. A., Life without a cerebellum. *Brain* **2010,** *133* (3), 652-654.
151. (a) Brahmachari, S.; Fung, Y. K.; Pahan, K., Induction of Glial Fibrillary Acidic Protein Expression in Astrocytes by Nitric Oxide. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **2006,** *26* (18), 4930-4939; (b) Ehrlich, S.; Burghardt, R.; Weiss, D.; Salbach-Andrae, H.; Craciun, E. M.; Goldhahn, K.; Klapp, B. F.; Lehmkuhl, U., Glial and neuronal damage markers in patients with anorexia nervosa. *Journal of neural transmission (Vienna, Austria : 1996)* **2008,** *115* (6), 921-7.
152. Alberts B, J. A., Lewis J, Raff M, Roberts K, Walter W, General Principles of Cell Communication. In *Molecular Biology of the Cell*, Garland Science: New York, 2002; Vol. 4th edition.
153. Saito, N.; Shirai, Y., Protein Kinase Cγ (PKCγ): Function of Neuron Specific Isotype. *The Journal of Biochemistry* **2002,** *132* (5), 683-687.
154. Bright, R.; Mochly-Rosen, D., The Role of Protein Kinase C in Cerebral Ischemic and Reperfusion Injury. *Stroke* **2005,** *36* (12), 2781-2790.
155. Dodge, F. A.; Rahamimoff, R., Co-operative action of calcium ions in transmitter release at the neuromuscular junction. *The Journal of Physiology* **1967,** *193* (2), 419-432.
156. Harris, J. K.; DeLorenzo, R. J., Calcium and neuronal cytoskeletal proteins: Alterations with aging. *Neurobiology of Aging* **1987,** *8* (4), 359-361.
157. Ohlendieck, K., Skeletal muscle proteomics: current approaches, technical challenges and emerging techniques. *Skeletal Muscle* **2011,** *1* (1), 1-15.
158. Kobayashi, T.; Takita, Y.; Suzuki, A.; Katsu, Y.; Iguchi, T.; Ohta, Y., Vacuolar Degeneration of Skeletal Muscle in Transgenic Mice Overexpressing ORP150. *Journal of Veterinary Medical Science* **2008,** *70* (1), 115-118.
159. Berchtold, M. W.; Brinkmeier, H.; Müntener, M., Calcium Ion in Skeletal Muscle: Its Crucial Role for Muscle Function, Plasticity, and Disease. *Physiological Reviews* **2000,** *80* (3), 1215-1265.
160. Beard, N. A.; Laver, D. R.; Dulhunty, A. F., Calsequestrin and the calcium release channel of skeletal and cardiac muscle. *Progress in Biophysics and Molecular Biology* **2004,** *85* (1), 33-69.
161. Nakahara, S.; Oka, N.; Raz, A., On the role of galectin-3 in cancer apoptosis. *Apoptosis* **2005,** *10* (2), 267-275.

162. Talapatra, S.; Wagner, J. D. O.; Thompson, C. B., Elongation factor-1 alpha is a selective regulator of growth factor withdrawal and ER stress-induced apoptosis. *Cell Death and Differentiation* **2002,** *9* (8), 856-861.
163. Park, D.; Xiang, A. P.; Mao, F. F.; Zhang, L.; Di, C.-G.; Liu, X.-M.; Shao, Y.; Ma, B.-F.; Lee, J.-H.; Ha, K.-S.; Walton, N.; Lahn, B. T., Nestin Is Required for the Proper Self-Renewal of Neural Stem Cells. *STEM CELLS* **2010,** *28* (12), 2162-2171.
164. Monnier, N.; Romero, N. B.; Lerale, J.; Nivoche, Y.; Qi, D.; MacLennan, D. H.; Fardeau, M.; Lunardi, J., An autosomal dominant congenital myopathy with cores and rods is associated with a neomutation in the RYR1 gene encoding the skeletal muscle ryanodine receptor. *Human Molecular Genetics* **2000,** *9* (18), 2599-2608.
165. Mkaouar-Rebai, E.; Ben Mahmoud, A.; Chamkha, I.; Chabchoub, I.; Kammoun, T.; Hachicha, M.; Fakhfakh, F., A novel MT-CO2 m.8249G > A pathogenic variation and the MT-TW m.5521G > A mutation in patients with mitochondrial myopathy. *Mitochondrial DNA* **2014,** *25* (5), 394-399.
166. Baird, M. F.; Graham, S. M.; Baker, J. S.; Bickerstaff, G. F., Creatine-Kinase- and Exercise-Related Muscle Damage Implications for Muscle Performance and Recovery. *Journal of Nutrition and Metabolism* **2012,** *2012*, 13.
167. Fischer, G.; Tradler, T.; Zarnt, T., The mode of action of peptidyl prolyl cis/trans isomerases in vivo: binding vs. catalysis. *FEBS Letters* **1998,** *426* (1), 17-20.
168. Roos, A.; Kollipara, L.; Buchkremer, S.; Labisch, T.; Brauers, E.; Gatz, C.; Lentz, C.; Gerardo-Nava, J.; Weis, J.; Zahedi, R. P., Cellular Signature of SIL1 Depletion: Disease Pathogenesis due to Alterations in Protein Composition Beyond the ER Machinery. *Mol Neurobiol* **2016,** *53* (8), 5527-41.
169. Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G., Gene Ontology: tool for the unification of biology. *Nat Genet* **2000,** *25* (1), 25-29.
170. Franceschini, A.; Szklarczyk, D.; Frankild, S.; Kuhn, M.; Simonovic, M.; Roth, A.; Lin, J.; Minguez, P.; Bork, P.; von Mering, C.; Jensen, L. J., STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* **2013,** *41* (Database issue), D808-15.
171. Palomo, G. M.; Manfredi, G., Exploring new pathways of neurodegeneration in ALS: The role of mitochondria quality control. *Brain research* **2014**.
172. Jiang, Y. L.; Ning, Y.; Ma, X. L.; Liu, Y. Y.; Wang, Y.; Zhang, Z.; Shan, C. X.; Xu, Y. D.; Yin, L. M.; Yang, Y. Q., Alteration of the proteome profile of the pancreas in diabetic rats induced by streptozotocin. *International journal of molecular medicine* **2011,** *28* (2), 153-60.
173. Kollipara, L.; Buchkremer, S.; Weis, J.; Brauers, E.; Hoss, M.; Rütten, S.; Caviedes, P.; Zahedi, R. P.; Roos, A., Proteome Profiling and Ultrastructural Characterization of the Human RCMH Cell Line: Myoblastic Properties and Suitability for Myopathological Studies. *Journal of Proteome Research* **2016,** *15* (3), 945-955.
174. Guo, X.; Trudgian, D. C.; Lemoff, A.; Yadavalli, S.; Mirzaei, H., Confetti: A Multiprotease Map of the HeLa Proteome for Comprehensive Proteomics. *Molecular & Cellular Proteomics* **2014,** *13* (6), 1573-1584.

References

175. Beck, M.; Schmidt, A.; Malmstroem, J.; Claassen, M.; Ori, A.; Szymborska, A.; Herzog, F.; Rinner, O.; Ellenberg, J.; Aebersold, R., The quantitative proteome of a human cell line. *Molecular Systems Biology* **2011,** *7* (1).
176. Vizcaino, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Rios, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; Binz, P.-A.; Xenarios, I.; Eisenacher, M.; Mayer, G.; Gatto, L.; Campos, A.; Chalkley, R. J.; Kraus, H.-J.; Albar, J. P.; Martinez-Bartolome, S.; Apweiler, R.; Omenn, G. S.; Martens, L.; Jones, A. R.; Hermjakob, H., ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotech* **2014,** *32* (3), 223-226.
177. Bauer, S.; Grossmann, S.; Vingron, M.; Robinson, P. N., Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* **2008,** *24* (14), 1650-1651.
178. Juretic, N.; Jorquera, G.; Caviedes, P.; Jaimovich, E.; Riveros, N., Electrical stimulation induces calcium-dependent up-regulation of neuregulin-1beta in dystrophic skeletal muscle cell lines. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* **2012,** *29* (5-6), 919-30.
179. Kislinger, T.; Gramolini, A. O.; Pan, Y.; Rahman, K.; MacLennan, D. H.; Emili, A., Proteome Dynamics during C2C12 Myoblast Differentiation. *Molecular & Cellular Proteomics* **2005,** *4* (7), 887-901.
180. Jaalouk, D. E.; Lammerding, J., Mechanotransduction gone awry. *Nat Rev Mol Cell Biol* **2009,** *10* (1), 63-73.
181. Blake, D. J.; Weir, A.; Newey, S. E.; Davies, K. E., Function and Genetics of Dystrophin and Dystrophin-Related Proteins in Muscle. *Physiological Reviews* **2002,** *82* (2), 291-329.
182. Devarajan, P.; Stabach, P. R.; Mann, A. S.; Ardito, T.; Kashgarian, M.; Morrow, J. S., Identification of a small cytoplasmic ankyrin (AnkG119) in the kidney and muscle that binds beta I sigma spectrin and associates with the Golgi apparatus. *The Journal of cell biology* **1996,** *133* (4), 819-30.
183. Lamb, G., Excitation–Contraction Coupling In Skeletal Muscle: Comparisons With Cardiac Muscle. *Clinical and Experimental Pharmacology and Physiology* **2000,** *27* (3), 216-224.
184. (a) Caviedes, R.; Liberona, J. L.; Hidalgo, J.; Tascon, S.; Salas, K.; Jaimovich, E., A human skeletal muscle cell line obtained from an adult donor. *Biochimica et biophysica acta* **1992,** *1134* (3), 247-55; (b) Liberona, J. L.; Caviedes, P.; Tascon, S.; Hidalgo, J.; Giglio, J. R.; Sampaio, S. V.; Caviedes, R.; Jaimovich, E., Expression of ion channels during differentiation of a human skeletal muscle cell line. *Journal of muscle research and cell motility* **1997,** *18* (5), 587-98; (c) Liberona, J. L.; Powell, J. A.; Shenoi, S.; Petherbridge, L.; Caviedes, R.; Jaimovich, E., Differences in both inositol 1,4,5-trisphosphate mass and inositol 1,4,5-trisphosphate receptors between normal and dystrophic skeletal muscle cell lines. *Muscle & nerve* **1998,** *21* (7), 902-9.
185. Goswami, A.; Jesse, C.; Chandrasekar, A.; Bushuven, E.; Vollrath, J.; Dreser, A.; Katona, I.; Beyer, C.; Johann, S.; Feller, A.; Grond, M.; Wagner, S.; Nikolin, S.; Troost, D.; Weis, J., Accumulation of STIM1 is associated with the degenerative muscle fibre phenotype in ALS and other neurogenic atrophies. *Neuropathology and applied neurobiology* **2014**.
186. (a) Gautel, M., Cytoskeletal protein kinases: titin and its relations in mechanosensing. *Pflugers Archiv : European journal of physiology* **2011,** *462* (1), 119-34; (b) Psatha, M. I.; Razi, M.; Koffer, A.; Moss, S. E.; Sacks, D. B.; Bolsover, S. R., Targeting of calcium:calmodulin signals to the cytoskeleton by IQGAP1. *Cell Calcium* **2007,** *41* (6), 593-605.

187. Desiere, F.; Deutsch, E.; Nesvizhskii, A.; Mallick, P.; King, N.; Eng, J.; Aderem, A.; Boyle, R.; Brunner, E.; Donohoe, S.; Fausto, N.; Hafen, E.; Hood, L.; Katze, M.; Kennedy, K.; Kregenow, F.; Lee, H.; Lin, B.; Martin, D.; Ranish, J.; Rawlings, D.; Samelson, L.; Shiio, Y.; Watts, J.; Wollscheid, B.; Wright, M.; Yan, W.; Yang, L.; Yi, E.; Zhang, H.; Aebersold, R., Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biology* **2004,** *6* (1), R9.

188. (a) Caron, M.; Imam-Sghiouar, N.; Poirier, F.; Le Caer, J. P.; Labas, V.; Joubert-Caron, R., Proteomic map and database of lymphoblastoid proteins. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2002,** *771* (1-2), 197-209; (b) Giuliano, S.; Agresta, A. M.; De Palma, A.; Viglio, S.; Mauri, P.; Fumagalli, M.; Iadarola, P.; Montalbetti, L.; Salvini, R.; Bardoni, A., Proteomic analysis of lymphoblastoid cells from Nasu-Hakola patients: a step forward in our understanding of this neurodegenerative disorder. *PLoS One* **2014,** *9* (12), e110073; (c) Lee, L. C.; Weng, Y. T.; Wu, Y. R.; Soong, B. W.; Tseng, Y. C.; Chen, C. M.; Lee-Chen, G. J., Downregulation of proteins involved in the endoplasmic reticulum stress response and Nrf2-ARE signaling in lymphoblastoid cells of spinocerebellar ataxia type 17. *Journal of neural transmission* **2014,** *121* (6), 601-10.

189. Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** *100* (12), 6940-6945.

190. (a) Cox, B.; Emili, A., Tissue subcellular fractionation and protein extraction for use in mass-spectrometry-based proteomics. *Nat. Protocols* **2006,** *1* (4), 1872-1878; (b) Huber, L. A.; Pfaller, K.; Vietor, I., Organelle Proteomics: Implications for Subcellular Fractionation in Proteomics. *Circulation research* **2003,** *92* (9), 962-968.

191. Warren, C. M.; Geenen, D. L.; Helseth, D. L.; Xu, H.; Solaro, R. J., Sub-Proteomic fractionation, iTRAQ, and OFFGEL-LC-MS/MS approaches to cardiac proteomics. *Journal of proteomics* **2010,** *73* (8), 1551-1561.

192. (a) Anttonen, A. K.; Siintola, E.; Tranebjaerg, L.; Iwata, N. K.; Bijlsma, E. K.; Meguro, H.; Ichikawa, Y.; Goto, J.; Kopra, O.; Lehesjoki, A. E., Novel SIL1 mutations and exclusion of functional candidate genes in Marinesco-Sjogren syndrome. *Eur J Hum Genet* **2008,** *16* (8), 961-9; (b) Noreau, A.; La Piana, R.; Marcoux, C.; Dion, P.; Brais, B.; Bernard, G.; Rouleau, G., Novel SIL1 mutations cause cerebellar ataxia and atrophy in a French-Canadian family. *Neurogenetics* **2015**, 1-4.

193. Méneret, A.; Wiame, E.; Marelli, C.; Lenglet, T.; Van Schaftingen, E.; Sedel, F., A serine synthesis defect presenting with a charcot-marie-tooth–like polyneuropathy. *Archives of Neurology* **2012,** *69* (7), 908-911.

194. Weitzmann, A.; Volkmer, J.; Zimmermann, R., The nucleotide exchange factor activity of Grp170 may explain the non-lethal phenotype of loss of Sil1 function in man and mouse. *FEBS Letters* **2006,** *580* (22), 5237-5240.

195. Ichhaporia, V. P.; Sanford, T.; Howes, J.; Marion, T. N.; Hendershot, L. M., Sil1, a nucleotide exchange factor for BiP, is not required for antibody assembly or secretion. *Molecular Biology of the Cell* **2015,** *26* (3), 420-429.

196. Rainbolt, T. K.; Saunders, J. M.; Wiseman, R. L., Stress-responsive regulation of mitochondria through the ER unfolded protein response. *Trends in endocrinology and metabolism: TEM* **2014,** *25* (10), 528-37.

References

197. Slemmer, J. E.; De Zeeuw, C. I.; Weber, J. T., Don't get too excited: mechanisms of glutamate-mediated Purkinje cell death. In *Progress in Brain Research*, Elsevier: 2005; Vol. Volume 148, pp 367-390.
198. (a) Meredith, C.; Herrmann, R.; Parry, C.; Liyanage, K.; Dye, D. E.; Durling, H. J.; Duff, R. M.; Beckman, K.; de Visser, M.; van der Graaff, M. M.; Hedera, P.; Fink, J. K.; Petty, E. M.; Lamont, P.; Fabian, V.; Bridges, L.; Voit, T.; Mastaglia, F. L.; Laing, N. G., Mutations in the Slow Skeletal Muscle Fiber Myosin Heavy Chain Gene (MYH7) Cause Laing Early-Onset Distal Myopathy (MPD1). *The American Journal of Human Genetics* **2004,** *75* (4), 703-708; (b) Viswanathan, M. C.; Kronert, W. A.; Melkani, G. C.; Cammarato, A.; Bernstein, S. I., Myosin Storage Myopathy Mutations Disrupt Myofibrillar Assembly/ Stability and Cause Progressive Muscle Degeneration in a Drosophila Model. *Biophysical Journal 106* (2), 777a.
199. Shaw, G.; Morse, S.; Ararat, M.; Graham, F. L., Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *The FASEB Journal* **2002,** *16* (8), 869-871.
200. Moha ou Maati, H.; Peyronnet, R.; Devader, C.; Veyssiere, J.; Labbal, F.; Gandin, C.; Mazella, J.; Heurteaux, C.; Borsotto, M., A Human TREK-1/HEK Cell Line: A Highly Efficient Screening Tool for Drug Development in Neurological Diseases. *PLoS ONE* **2011,** *6* (10), e25602.
201. Schlachetzki, J. C. M.; Saliba, S. W.; Oliveira, A. C. P. d., Studying neurodegenerative diseases in culture models. *Revista Brasileira de Psiquiatria* **2013,** *35*, S92-S100.

7 Acknowledgements

I express my deep and sincere gratitude to **Prof. Dr. Albert Sickmann**, **Dr. René Peiman Zahedi** and **Dr. Andreas Roos** for their constant supervision, support and guidance during my doctoral studies. I also would like to thank **Prof. Dr. Albert Sickmann**, **Prof. Dr. Oliver Kayser** and **Prof. Dr. Markus Nett** for being part of my PhD thesis committee.

I gratefully acknowledge the financial support by the German Research Foundation - Deutsche Forschungsgemeinschaft (DFG to R. Z.; grant no. ZA 639/1-1).

I convey my gratitude to all the MSS patient families, the index patient and respective healthy volunteers who were studied in this work and underwent biopsy procedure. Unfortunately, the female index patient died in the year 2013 and I express my condolences to the family members of the deceased.

My sincere thanks to **Dr. Andreas Roos** and his colleagues at the Institute of Neuropathology, Aachen, for preparing and providing MSS-related samples to carry out my research activities at the Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V., Dortmund.

I would like to acknowledge the work of **BSc. Jennifer Baumann**, which helped me to choose a comparatively better off-line peptide fractionation strategy for my experiments.

It gives me a great pleasure to thank all my past and present co-workers from different research groups and departments at ISAS for providing a friendly and stimulating scientific environment.

Finally, I would like to thank my parents and family members for their unconditional affection and support.

8 Erklärung

Hiermit erkläre ich an Eides statt, dass ich die Dissertation „Characterizing protein processing in the Endoplasmic Reticulum using quantitative proteomics: the pathogenesis of the Marinesco-Sjörgren Syndrome“ selbständig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe.

Ich erkläre außerdem, dass diese Dissertation weder in gleicher oder anderer Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ich habe früher außer den mit dem Zulassungsgesuch urkundlich vorgelegten Graden keine weiteren akademischen Grade erworben oder zu erwerben versucht.

Dortmund, den

9 Curriculum vitae

First name:	Laxmikanth
Family name:	Kollipara
Marital status:	Single
Email ID:	laxmikanth.kollipara@isas.de
Telephone:	(+49)02311392-4220 (O) / 01783633519 (H)
Address:	Adalbertstr. 149, Apartment 103 44149, Dortmund
Date of birth:	24.07.1981
Place of birth:	Hyderabad, India

Education:

Dec-2010 to Dec-2016: PhD dissertation at the Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V., Dortmund, Germany affiliated to Technical University of Dortmund, Germany.

Oct-2008 to Dec-2010: Master of Science (MSc.) in Pharmaceutical Biotechnology (120 ECTS credits) from Martin Luther University of Halle-Wittenberg, Halle, Germany.

Nov-2007 to Sep-2008: Master of Science in Applied Polymer Science from Martin Luther University of Halle-Wittenberg, Halle, Germany (discontinued).

Jan-2000 to Apr-2004: Bachelor of Pharmacy (B. Pharm) from Jawahar Lal Nehru Technology University, India.

Work experience:

Jan-2014 till date: Research associate at the Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V., Dortmund, Germany.

Jun-2005 to Oct-2007: Independent study and training.

Sep-2004 to May-2005: Business executive at the Emcure Pharmaceuticals Ltd., Pune, India.

10 Appendices

10.1 Instruments and LC-MS parameters used - CID fragmentation

Instrument/Setup	Settings/Parameters	Experiment name		
		Protein carbamylation (Peptide mixtures)	**Protein carbamylation (2-step digest, Fibroblasts)**	**Human myoblastic RCMH; label-free**
LC-MS		U3000 nRSLC-LTQ Orbitrap XL	U3000 HPLC-LTQ Orbitrap Velos	U3000 nRSLC-Orbitrap Elite
HPLC	Column type, length	Commercial, 15 cm	Commercial, 50 cm	Commercial, 50 cm
	Gradient	5-50% B in 50 min	3-42% B in 120 min	3-42% B in 187 min
MS	Polarity	Positive	Positive	Positive
	Data acquisition mode, Top N	DDA, Top 5	DDA, Top 15	DDA, Top 15
	Scan range (m/z)	300 - 2,000	300 - 1,500	300 - 1,500
	Resolution (FWHM)			
	Full MS scan (Orbitrap)	60,000 at m/z 400	60,000 at m/z 400	60,000 at m/z 400
	MS/MS scan (Ion trap)	0.7 u	0.45 u	0.45 u
	Automated gain control (AGC) target values			
	Full MS	5×10^5	1×10^6	1×10^6
	MS/MS	1×10^4	1×10^4	1×10^4
	Maximum ion injection times (ms)			
	Full MS	500	100	100
	MS/MS	200	100	100
	Fragmentation type	CID	CID	CID
	Precursor isolation width (m/z)	2.0	2.0	2.0
	Dynamic exclusion duration (s)	10	30	30
	Normalized collision energy (%)	35	35	35
	Activation time (ms)	30	10	10
	Lock mass - polysiloxane ion (m/z)	371.101236	371.101236	371.101236

10.2 Instruments and LC-MS parameters used - HCD fragmentation

Instrument/Setup	**Settings/Parameters**	**Experiment name**	
		Human EBV-LCs and Mice cerebella; iTRAQ 8-plex	**Human skeletal muscles; label-free**
LC-MS		U3000 HPLC-LTQ Orbitrap Velos	U3000 nRSLC-Q Exactive
HPLC	Column type, length	Self-packed, 30 cm	Commercial, 50 cm
	Gradient	3-45% B in 120 min	3-35% B in 187 min
MS	Polarity	Positive	Positive
	Data acquisition mode, Top N	DDA, Top 5	DDA, Top 15
	Scan range (m/z)	300 - 2,000	300 - 1,500
	Resolution (FWHM)		
	Full MS (Orbitrap)	30,000 at m/z 400	70,000 at m/z 200
	MS/MS (Orbitrap)	7,500 at m/z 400	17,500 at m/z 200
	Automated gain control (AGC) target values		
	Full MS	1×10^6	2×10^5
	MS/MS	1×10^5	5×10^4
	Maximum ion injection times (ms)		
	Full MS	100	120
	MS/MS	200	250
	Fragmentation type	HCD	HCD
	Precursor isolation width (m/z)	2.0	2.0
	Dynamic exclusion duration (s)	30	12
	Normalized collision energy (%)	47	27
	Activation time (ms)	0.2	Not applicable
	Underfill ratio	Not applicable	5%
	Lock mass - polysiloxane ion (m/z)	371.101236	371.101236

10.3 Significantly altered proteins in the MSS-fibroblasts study

UniProt accession	Protein	Gene	Unique Peptides	PSMs	MSS/Healthy	RSD%
P05120	Plasminogen activator inhibitor 2	*SERPINB2*	21	308	12.1	4%
Q8TF66	Leucine-rich repeat-containing protein 15	*LRRC15*	5	24	8.8	4%
P07197	Neurofilament medium polypeptide	*NEFM*	3	162	7.7	5%
Q6P179	Endoplasmic reticulum aminopeptidase 2	*ERAP2*	3	15	7.2	10%
O00469	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	*PLOD2*	23	209	7.0	5%
P28300	Protein-lysine 6-oxidase	*LOX*	4	20	6.7	6%
P08254	Stromelysin-1	*MMP3*	12	126	6.4	12%
P03956	Interstitial collagenase	*MMP1*	15	166	5.9	5%
P07996	Thrombospondin-1	*THBS1*	21	147	5.8	3%
P05362	Intercellular adhesion molecule 1	*ICAM1*	5	24	5.8	3%
P04179	Superoxide dismutase [Mn], mitochondrial	*SOD2*	9	193	5.1	4%
P01584	Interleukin-1 beta	*IL1B*	4	25	4.8	3%
P04439	HLA class I histocompatibility antigen, A-3 alpha chain	*HLA-A*	2	57	4.5	10%
Q14627	Interleukin-13 receptor subunit alpha-2	*IL13RA2*	2	2	4.4	8%
P07951	Tropomyosin beta chain	*TPM2*	5	203	4.2	14%
P04004	Vitronectin	*VTN*	2	5	4.1	8%
P05121	Plasminogen activator inhibitor 1	*SERPINE1*	12	119	4.1	6%
Q06033	Inter-alpha-trypsin inhibitor heavy chain H3	*ITIH3*	7	29	4.0	3%
P12277	Creatine kinase B-type	*CKB*	8	36	3.9	3%
P20936	Ras GTPase-activating protein 1	*RASA1*	16	166	3.9	3%
O14558	Heat shock protein beta-6	*HSPB6*	5	65	3.8	1%
Q14624	Inter-alpha-trypsin inhibitor heavy chain H4	*ITIH4*	2	16	3.5	6%
O43719	HIV Tat-specific factor 1	*HTATSF1*	3	8	3.4	9%
P09493	Tropomyosin alpha-1 chain	*TPM1*	7	136	3.4	8%
O15427	Monocarboxylate transporter 4	*SLC16A3*	6	27	3.3	2%
Q01995	Transgelin	*TAGLN*	13	329	3.3	2%
P07093	Glia-derived nexin	*SERPINE2*	10	130	3.2	4%

P01023	Alpha-2-macroglobulin	*A2M*	5	170	3.2	3%
O15460	Prolyl 4-hydroxylase subunit alpha-2	*P4HA2*	13	74	3.1	5%
P55290	Cadherin-13	*CDH13*	2	3	3.1	2%
P09936	Ubiquitin carboxyl-terminal hydrolase isozyme L1	*UCHL1*	10	150	3.1	1%
P24821	Tenascin	*TNC*	17	71	2.9	3%
Q96D15	Reticulocalbin-3	*RCN3*	8	82	2.7	7%
P50479	PDZ and LIM domain protein 4	*PDLIM4*	6	13	2.6	1%
P09038	Fibroblast growth factor 2	*FGF2*	3	19	2.6	5%
P09601	Heme oxygenase 1	*HMOX1*	10	56	2.6	8%
P10253	Lysosomal alpha-glucosidase	*GAA*	4	15	2.5	3%
Q15582	Transforming growth factor-beta-induced protein ig-h3	*TGFBI*	4	13	2.5	2%
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	*HSPG2*	17	51	2.5	3%
Q96S97	Myeloid-associated differentiation marker	*MYADM*	2	70	2.5	3%
P09486	SPARC	*SPARC*	5	12	2.5	6%
P07585	Decorin	*DCN*	3	14	2.4	18%
Q32MZ4	Leucine-rich repeat flightless-interacting protein 1	*LRRFIP1*	6	10	2.4	5%
Q99439	Calponin-2	*CNN2*	8	110	2.4	1%
O95340	Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 2	*PAPSS2*	10	41	2.4	2%
P84157	Matrix-remodeling-associated protein 7	*MXRA7*	2	21	2.4	2%
P04114	Apolipoprotein B-100	*APOB*	8	29	2.3	4%
Q96NE9	FERM domain-containing protein 6	*FRMD6*	2	6	2.3	8%
P98082	Disabled homolog 2	*DAB2*	13	46	2.3	6%
P20337	Ras-related protein Rab-3B	*RAB3B*	5	62	2.3	12%
Q8NBZ7	UDP-glucuronic acid decarboxylase 1	*UXS1*	2	4	2.3	14%
Q9UHB6	LIM domain and actin-binding protein 1	*LIMA1*	15	85	2.3	2%
P19971	Thymidine phosphorylase	*TYMP*	3	26	2.3	5%
P01033	Metalloproteinase inhibitor 1	*TIMP1*	2	5	2.3	6%
Q07065	Cytoskeleton-associated protein 4	*CKAP4*	39	777	2.2	4%
P11233	Ras-related protein Ral-A	*RALA*	2	15	2.2	12%
P69905	Hemoglobin subunit alpha	*HBA1*	5	86	2.2	13%
Q9UBG0	C-type mannose receptor 2	*MRC2*	11	80	2.2	2%
P11166	Solute carrier family 2, facilitated glucose transporter member 1	*SLC2A1*	3	29	2.2	9%
P02649	Apolipoprotein E	*APOE*	2	6	2.2	15%
P52848	Bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 1	*NDST1*	2	4	2.2	4%

Q96PD2	Discoidin, CUB and LCCL domain-containing protein 2	*DCBLD2*	2	10	2.2	4%
Q15274	Nicotinate-nucleotide pyrophosphorylase [carboxylating]	*QPRT*	4	19	2.2	6%
P13674	Prolyl 4-hydroxylase subunit alpha-1	*P4HA1*	19	114	2.1	3%
Q96D46	60S ribosomal export protein NMD3	*NMD3*	2	5	2.1	20%
P05161	Ubiquitin-like protein ISG15	*ISG15*	3	45	2.1	5%
P43490	Nicotinamide phosphoribosyltransferase	*NAMPT*	18	219	2.1	1%
P02511	Alpha-crystallin B chain	*CRYAB*	6	37	2.1	3%
Q9UJ70	N-acetyl-D-glucosamine kinase	*NAGK*	10	53	2.1	4%
P02765	Alpha-2-HS-glycoprotein	*AHSG*	4	55	2.1	10%
P35579	Myosin-9	*MYH9*	115	2114	2.1	3%
Q9UBR2	Cathepsin Z	*CTSZ*	4	56	2.1	3%
O60711	Leupaxin	*LPXN*	6	28	2.1	7%
P17936	Insulin-like growth factor-binding protein 3	*IGFBP3*	2	8	2.1	7%
P51805	Plexin-A3	*PLXNA3*	2	7	2.1	14%
P04216	Thy-1 membrane glycoprotein	*THY1*	4	51	2.1	2%
Q0ZGT2	Nexilin	*NEXN*	4	5	2.1	2%
Q9NR12	PDZ and LIM domain protein 7	*PDLIM7*	7	25	2.1	7%
P52789	Hexokinase-2	*HK2*	4	121	2.0	3%
O43175	D-3-phosphoglycerate dehydrogenase	*PHGDH*	12	96	0.5	5%
P30837	Aldehyde dehydrogenase X, mitochondrial	*ALDH1B1*	5	34	0.5	3%
Q12791	Calcium-activated potassium channel subunit alpha-1	*KCNMA1*	2	6	0.5	2%
P29372	DNA-3-methyladenine glycosylase	*MPG*	3	19	0.5	8%
Q96PY5	Formin-like protein 2	*FMNL2*	6	46	0.5	3%
Q5SSJ5	Heterochromatin protein 1-binding protein 3	*HP1BP3*	10	48	0.5	3%
Q6ZMZ3	Nesprin-3	*C14orf49*	5	9	0.5	7%
Q8NEY1	Neuron navigator 1	*NAV1*	5	14	0.5	7%
O75828	Carbonyl reductase [NADPH] 3	*CBR3*	3	140	0.5	2%
Q9BYC5	Alpha-(1,6)-fucosyltransferase	*FUT8*	2	3	0.5	8%
P16401	Histone H1.5	*HIST1H1B*	7	145	0.5	12%
Q9BTZ2	Dehydrogenase/reductase SDR family member 4	*DHRS4*	4	45	0.5	4%
P57764	Gasdermin-D	*GSDMD*	3	8	0.5	13%
Q8WWI5	Choline transporter-like protein 1	*SLC44A1*	5	35	0.5	4%
Q92466	DNA damage-binding protein 2	*DDB2*	2	3	0.5	12%
Q5KU26	Collectin-12	*COLEC12*	13	48	0.5	2%

P20700	Lamin-B1	*LMNB1*	23	164	0.5	5%
Q14699	Raftlin	*RFTN1*	10	29	0.5	3%
P11498	Pyruvate carboxylase, mitochondrial	*PC*	7	33	0.5	7%
Q9BQ39	ATP-dependent RNA helicase DDX50	*DDX50*	2	6	0.5	8%
Q5T9L3	Protein wntless homolog	*WLS*	5	27	0.5	5%
P14384	Carboxypeptidase M	*CPM*	2	5	0.4	7%
P62995	Transformer-2 protein homolog beta	*TRA2B*	2	6	0.4	5%
Q16719	Kynureninase	*KYNU*	3	8	0.4	7%
Q8N4T8	Carbonyl reductase family member 4	*CBR4*	2	12	0.4	0%
Q7L5Y1	Mitochondrial enolase superfamily member 1	*ENOSF1*	2	3	0.4	14%
P17096	High mobility group protein HMG-I/HMG-Y	*HMGA1*	2	11	0.4	3%
Q16666	Gamma-interferon-inducible protein 16	*IFI16*	14	57	0.4	7%
Q9UDY2	Tight junction protein ZO-2	*TJP2*	4	10	0.4	7%
O14684	Prostaglandin E synthase	*PTGES*	3	18	0.4	8%
Q8N2H3	Pyridine nucleotide-disulfide oxidoreductase domain-containing protein 2	*PYROXD2*	5	16	0.4	6%
Q92506	Estradiol 17-beta-dehydrogenase 8	*HSD17B8*	4	24	0.4	6%
P51648	Fatty aldehyde dehydrogenase	*ALDH3A2*	12	46	0.4	4%
Q96CX2	BTB/POZ domain-containing protein KCTD12	*KCTD12*	12	228	0.4	5%
Q9HA77	Probable cysteine--tRNA ligase, mitochondrial	*CARS2*	2	4	0.4	11%
P47712	Cytosolic phospholipase A2	*PLA2G4A*	6	44	0.4	4%
Q9H173	Nucleotide exchange factor SIL1	*SIL1*	2	2	0.4	10%
Q53EL6	Programmed cell death protein 4	*PDCD4*	5	49	0.4	8%
Q9Y3Z3	SAM domain and HD domain-containing protein 1	*SAMHD1*	4	7	0.3	12%
P16402	Histone H1.3	*HIST1H1D*	2	188	0.3	13%
P62805	Histone H4	*HIST1H4A*	9	600	0.3	5%
P52895	Aldo-keto reductase family 1 member C2	*AKR1C2*	3	233	0.3	3%
P21266	Glutathione S-transferase Mu 3	*GSTM3*	5	58	0.3	8%
Q9Y4K1	Absent in melanoma 1 protein	*AIM1*	2	6	0.3	10%
Q658P3	Metalloreductase STEAP3	*STEAP3*	7	77	0.3	9%
Q16647	Prostacyclin synthase	*PTGIS*	5	13	0.3	9%
Q13228	Selenium-binding protein 1	*SELENBP1*	10	84	0.3	6%
P08294	Extracellular superoxide dismutase [Cu-Zn]	*SOD3*	3	19	0.3	10%
Q9P2B2	Prostaglandin F2 receptor negative regulator	*PTGFRN*	3	7	0.3	13%
Q05707	Collagen alpha-1(XIV) chain	*COL14A1*	2	15	0.2	17%

P00966	Argininosuccinate synthase	*ASS1*	19	261	0.2	4%
Q5BJF2	Transmembrane protein 97	*TMEM97*	2	3	0.2	13%
O00534	von Willebrand factor A domain-containing protein 5A	*VWA5A*	2	3	0.2	11%
O95425	Supervillin	*SVIL*	14	58	0.2	5%
O60437	Periplakin	*PPL*	31	122	0.2	10%
O14495	Lipid phosphate phosphohydrolase 3	*PPAP2B*	5	58	0.2	13%
P29762	Cellular retinoic acid-binding protein 1	*CRABP1*	3	62	0.1	17%

10.4 Significantly altered proteins in the MSS-LCs study

UniProt accession	Protein	Gene	Unique Peptides	PSMs	MSS/Healthy	T.TEST
Q96AC1	Fermitin family homolog 2	*FERMT2*	4	10	3.7	0.00
P08631	Tyrosine-protein kinase HCK	*HCK*	5	29	2.8	0.05
P35080	Profilin-2	*PFN2*	4	7	2.6	0.03
P04066	Tissue alpha-L-fucosidase	*FUCA1*	6	14	2.2	0.00
Q8N584	Tetratricopeptide repeat protein 39C	*TTC39C*	3	5	2.0	0.04
Q14108	Lysosome membrane protein 2	*SCARB2*	2	5	1.8	0.02
P28068	HLA class II histocompatibility antigen, DM beta chain	*HLA-DMB*	4	9	1.8	0.01
P06239	Tyrosine-protein kinase Lck	*LCK*	12	49	1.8	0.02
P09917	Arachidonate 5-lipoxygenase	*ALOX5*	12	41	1.7	0.03
P06340	HLA class II histocompatibility antigen, DO alpha chain	*HLA-DOA*	2	5	1.7	0.02
Q7Z3E5	LisH domain-containing protein ARMC9	*ARMC9*	4	10	1.7	0.01
Q7Z4S6	Kinesin-like protein KIF21A	*KIF21A*	9	27	1.7	0.04
P07203	Glutathione peroxidase 1	*GPX1*	9	37	1.7	0.01
P04114	Apolipoprotein B-100	*APOB*	7	18	1.7	0.01
P28067	HLA class II histocompatibility antigen, DM alpha chain	*HLA-DMA*	2	5	1.6	0.00
Q14005	Pro-interleukin-16	*IL16*	14	51	1.6	0.02
O95671	N-acetylserotonin O-methyltransferase-like protein	*ASMTL*	19	79	1.5	0.03
P02774	Vitamin D-binding protein	*GC*	4	6	1.5	0.01
Q86W92	Liprin-beta-1	*PPFIBP1*	2	3	1.5	0.02
P36969	Phospholipid hydroperoxide glutathione peroxidase, mitochondrial	*GPX4*	7	22	1.5	0.04
O95571	Persulfide dioxygenase ETHE1, mitochondrial	*ETHE1*	9	31	1.5	0.02
Q8IY21	Probable ATP-dependent RNA helicase DDX60	*DDX60*	15	41	1.5	0.01

Q6WKZ4	Rab11 family-interacting protein 1	*RAB11FIP1*	9	21	1.5	0.02
Q9UBB4	Ataxin-10	*ATXN10*	19	69	1.5	0.01
P19823	Inter-alpha-trypsin inhibitor heavy chain H2	*ITIH2*	4	12	1.5	0.04
P07858	Cathepsin B	*CTSB*	2	5	1.5	0.00
Q8IZQ5	Selenoprotein H	*SELH*	3	7	1.5	0.00
Q8WYJ6	Septin-1	*Sep 01*	18	87	1.5	0.01
P16219	Short-chain specific acyl-CoA dehydrogenase, mitochondrial	*ACADS*	9	30	1.5	0.02
Q9BRQ8	Apoptosis-inducing factor 2	*AIFM2*	2	3	1.4	0.04
P11310	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial	*ACADM*	19	122	1.4	0.04
P04062	Glucosylceramidase	*GBA*	2	3	1.4	0.00
Q8IZ07	Ankyrin repeat domain-containing protein 13A	*ANKRD13A*	9	33	1.4	0.05
P00395	Cytochrome c oxidase subunit 1	*MT-CO1*	3	21	1.4	0.01
P13765	HLA class II histocompatibility antigen, DO beta chain	*HLA-DOB*	11	20	1.4	0.01
Q5R372	Rab GTPase-activating protein 1-like	*RABGAP1L*	20	67	1.4	0.01
Q9UPR0	Inactive phospholipase C-like protein 2	*PLCL2*	3	5	1.4	0.01
Q9P1Z2	Calcium-binding and coiled-coil domain-containing protein 1	*CALCOCO1*	2	4	1.4	0.01
P41218	Myeloid cell nuclear differentiation antigen	*MNDA*	20	44	1.4	0.04
Q96LZ7	Regulator of microtubule dynamics protein 2	*RMDN2*	2	4	1.4	0.00
Q8WZA0	Protein LZIC	*LZIC*	3	9	1.4	0.05
Q14699	Raftlin	*RFTN1*	14	53	1.4	0.02
P59768	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-2	*GNG2*	4	7	1.4	0.04
P19256	Lymphocyte function-associated antigen 3	*CD58*	3	8	1.4	0.00
Q9BW62	Katanin p60 ATPase-containing subunit A-like 1	*KATNAL1*	2	5	1.4	0.02
P02751	Fibronectin	*FN1*	6	11	1.4	0.00
Q9BRX8	Redox-regulatory protein FAM213A	*FAM213A*	5	16	1.4	0.01
Q12965	Unconventional myosin-Ie	*MYO1E*	27	100	1.4	0.02
P30044	Peroxiredoxin-5, mitochondrial	*PRDX5*	10	82	1.4	0.01
P99999	Cytochrome c	*CYCS*	7	50	1.4	0.01
P09382	Galectin-1	*LGALS1*	13	108	1.4	0.04
P45954	Short/branched chain specific acyl-CoA dehydrogenase, mitochondrial	*ACADSB*	9	26	1.4	0.02
P35625	Metalloproteinase inhibitor 3	*TIMP3*	3	6	1.3	0.00
Q9NXH8	Torsin-4A	*TOR4A*	4	9	1.3	0.02
O00750	Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit beta	*PIK3C2B*	3	5	1.3	0.01
P49641	Alpha-mannosidase 2x	*MAN2A2*	4	13	1.3	0.01
Q9UHG3	Prenylcysteine oxidase 1	*PCYOX1*	7	19	1.3	0.02
O60711	Leupaxin	*LPXN*	14	45	1.3	0.04

O75695	Protein XRP2	*RP2*	5	7	1.3	0.04
Q6P2P2	Putative protein arginine N-methyltransferase 10	*PRMT10*	3	8	1.3	0.00
Q8IX04	Ubiquitin-conjugating enzyme E2 variant 3	*UEVLD*	3	7	1.3	0.01
Q08722	Leukocyte surface antigen CD47	*CD47*	2	10	1.3	0.02
Q05823	2-5A-dependent ribonuclease	*RNASEL*	2	4	1.3	0.01
Q9HBL8	NmrA-like family domain-containing protein 1	*NMRAL1*	6	18	1.3	0.04
Q96S94	Cyclin-L2	*CCNL2*	2	3	0.8	0.04
Q9H9Y2	Ribosome production factor 1	*RPF1*	2	3	0.8	0.03
P13674	Prolyl 4-hydroxylase subunit alpha-1	*P4HA1*	17	48	0.8	0.03
Q9BVV7	Mitochondrial import inner membrane translocase subunit Tim21	*TIMM21*	2	2	0.8	0.01
Q6ZVM7	TOM1-like protein 2	*TOM1L2*	2	2	0.8	0.03
O00165	HCLS1-associated protein X-1	*HAX1*	4	4	0.8	0.01
Q6N063	2-oxoglutarate and iron-dependent oxygenase domain-containing protein 2	*OGFOD2*	2	5	0.8	0.01
O75794	Cell division cycle protein 123 homolog	*CDC123*	3	7	0.7	0.00
Q13488	V-type proton ATPase 116 kDa subunit a isoform 3	*TCIRG1*	3	5	0.7	0.00
Q8TBF2	Prostamide/prostaglandin F synthase	*FAM213B*	2	3	0.7	0.03
Q00534	Cyclin-dependent kinase 6	*CDK6*	10	27	0.7	0.05
Q15032	R3H domain-containing protein 1	*R3HDM1*	2	3	0.7	0.05
Q9Y617	Phosphoserine aminotransferase	*PSAT1*	13	47	0.7	0.01
Q9NNW7	Thioredoxin reductase 2, mitochondrial	*TXNRD2*	4	13	0.7	0.01
Q4VC31	Coiled-coil domain-containing protein 58	*CCDC58*	2	6	0.7	0.01
P18085	ADP-ribosylation factor 4	*ARF4*	7	75	0.7	0.01
Q14213	Interleukin-27 subunit beta	*EBI3*	6	17	0.7	0.01
Q9Y6E2	Basic leucine zipper and W2 domain-containing protein 2	*BZW2*	4	14	0.7	0.02
Q7Z417	Nuclear fragile X mental retardation-interacting protein 2	*NUFIP2*	2	3	0.7	0.01
P35610	Sterol O-acyltransferase 1	*SOAT1*	2	4	0.7	0.01
Q96DU3	SLAM family member 6	*SLAMF6*	2	3	0.7	0.01
Q8NFC6	Biorientation of chromosomes in cell division protein 1-like 1	*BOD1L1*	2	5	0.7	0.03
Q15424	Scaffold attachment factor B1	*SAFB*	14	72	0.7	0.02
P63218	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-5	*GNG5*	2	8	0.7	0.03
Q16222	UDP-N-acetylhexosamine pyrophosphorylase	*UAP1*	6	19	0.7	0.03
Q9NYL4	Peptidyl-prolyl cis-trans isomerase FKBP11	*FKBP11*	4	10	0.7	0.03
Q8WV28	B-cell linker protein	*BLNK*	9	15	0.7	0.02
P29966	Myristoylated alanine-rich C-kinase substrate	*MARCKS*	3	8	0.7	0.01
P84101	Small EDRK-rich factor 2	*SERF2*	3	11	0.7	0.01
Q70UQ0	Inhibitor of nuclear factor kappa-B kinase-interacting protein	*IKBIP*	3	5	0.7	0.04

P62158	Calmodulin	*CALM1*	4	16	0.7	0.05
Q86V48	Leucine zipper protein 1	*LUZP1*	3	4	0.7	0.03
P51397	Death-associated protein 1	*DAP*	2	5	0.7	0.01
Q9UGC7	Peptide chain release factor 1-like, mitochondrial	*MTRF1L*	2	4	0.6	0.01
Q6UXH1	Cysteine-rich with EGF-like domain protein 2	*CRELD2*	2	7	0.6	0.05
P06737	Glycogen phosphorylase, liver form	*PYGL*	4	18	0.6	0.00
P02794	Ferritin heavy chain	*FTH1*	2	3	0.6	0.05
P42331	Rho GTPase-activating protein 25	*ARHGAP25*	4	15	0.6	0.04
P08236	Beta-glucuronidase	*GUSB*	2	3	0.6	0.00
P28908	Tumor necrosis factor receptor superfamily member 8	*TNFRSF8*	3	6	0.6	0.00
P50416	Carnitine O-palmitoyltransferase 1, liver isoform	*CPT1A*	13	30	0.6	0.01
Q9BX59	Tapasin-related protein	*TAPBPL*	2	3	0.6	0.00
P06454	Prothymosin alpha	*PTMA*	3	10	0.6	0.04
O43175	D-3-phosphoglycerate dehydrogenase	*PHGDH*	17	64	0.5	0.03
P02792	Ferritin light chain	*FTL*	4	5	0.4	0.01
P04181	Ornithine aminotransferase, mitochondrial	*OAT*	3	7	0.4	0.03
P21266	Glutathione S-transferase Mu 3	*GSTM3*	5	18	0.4	0.02
Q9H173	Nucleotide exchange factor SIL1	*SIL1*	3	10	0.3	0.00

10.5 Significantly altered proteins in the *woozy* mice cerebella study

UniProt accession	Protein	Gene	Unique Peptides	PSMs	*woozy*/wild type	T.TEST
P03995	Glial fibrillary acidic protein	*Gfap*	35	191	1.9	0.01
P70677	Caspase-3	*Casp3*	2	4	1.7	0.00
Q91XV3	Brain acid soluble protein 1	*Basp1*	8	24	1.6	0.05
P23242	Gap junction alpha-1 protein	*Gja1*	8	29	1.5	0.04
Q9WUC3	Lymphocyte antigen 6H	*Ly6h*	3	6	1.5	0.05
Q8BWU8	Ethanolamine-phosphate phospho-lyase	*Etnppl*	9	29	1.5	0.01
O88533	Aromatic-L-amino-acid decarboxylase	*Ddc*	4	10	1.5	0.04
P61458	Pterin-4-alpha-carbinolamine dehydratase	*Pcbd1*	4	8	1.5	0.04
Q99P58	Ras-related protein Rab-27B	*Rab27b*	4	9	1.5	0.03
P11859	Angiotensinogen	*Agt*	2	5	1.5	0.00
P55088	Aquaporin-4	*Aqp4*	4	11	1.4	0.00
O88492	Perilipin-4	*Plin4*	4	7	1.4	0.01

P31650	Sodium- and chloride-dependent GABA transporter 3	*Slc6a11*	9	30	1.4	0.03
P56528	ADP-ribosyl cyclase 1	*Cd38*	3	4	1.4	0.01
Q8BZF8	Phosphoglucomutase-like protein 5	*Pgm5*	2	4	1.4	0.00
O08709	Peroxiredoxin-6	*Prdx6*	20	143	1.4	0.03
Q8BW75	Amine oxidase [flavin-containing] B	*Maob*	14	42	1.4	0.00
P70689	Gap junction beta-6 protein	*Gjb6*	2	6	1.4	0.00
P05532	Mast/stem cell growth factor receptor Kit	*Kit*	4	9	1.4	0.04
Q3TVI8	Pre-B-cell leukemia transcription factor-interacting protein 1	*Pbxip1*	5	13	1.4	0.01
Q3UNZ8	Quinone oxidoreductase-like protein 2	*N/A*	2	3	1.4	0.00
P20152	Vimentin	*Vim*	31	145	1.4	0.01
Q62421	Endophilin-A3	*Sh3gl3*	4	7	1.4	0.04
Q8CC86	Nicotinate phosphoribosyltransferase	*Naprt1*	5	6	1.3	0.00
Q80XD1	Beta-chimaerin	*Chn2*	2	6	0.7	0.02
Q8VHC3	Selenoprotein M	*Selm*	4	17	0.7	0.00
Q63959	Potassium voltage-gated channel subfamily C member 3	*Kcnc3*	9	36	0.7	0.00
Q6NVG1	Lysophospholipid acyltransferase LPCAT4	*Lpcat4*	7	15	0.7	0.00
P13707	Glycerol-3-phosphate dehydrogenase [NAD(+)], cytoplasmic	*Gpd1*	22	134	0.7	0.01
Q8C1B7	Septin-11	*Sep 11*	4	73	0.7	0.00
Q3UUJ4	STE20-related kinase adapter protein alpha	*Strada*	2	2	0.7	0.04
Q8BYR5	Calcium-dependent secretion activator 2	*Cadps2*	16	58	0.7	0.02
P59644	Phosphatidylinositol 4,5-bisphosphate 5-phosphatase A	*Inpp5j*	4	10	0.7	0.02
Q91Z69	SLIT-ROBO Rho GTPase-activating protein 1	*Srgap1*	2	6	0.7	0.02
Q60598	Src substrate cortactin	*Cttn*	11	31	0.7	0.00
Q9R0K7	Plasma membrane calcium-transporting ATPase 2	*Atp2b2*	53	193	0.7	0.00
P28659	CUGBP Elav-like family member 1	*Celf1*	2	4	0.7	0.01
Q80T41	Gamma-aminobutyric acid type B receptor subunit 2	*Gabbr2*	12	28	0.7	0.01
A2ANU3	Synapse differentiation-inducing gene protein 1	*Syndig1*	2	3	0.7	0.00
P97450	ATP synthase-coupling factor 6, mitochondrial	*Atp5j*	2	7	0.7	0.03
Q91WG7	Diacylglycerol kinase gamma	*Dgkg*	5	9	0.7	0.02
P28652	Calcium/calmodulin-dependent protein kinase type II subunit beta	*Camk2b*	11	75	0.7	0.02
Q6PHS9	Voltage-dependent calcium channel subunit alpha-2/delta-2	*Cacna2d2*	16	42	0.7	0.00
Q1RLL3	Copine-9	*Cpne9*	6	20	0.7	0.02
P51880	Fatty acid-binding protein, brain	*Fabp7*	3	17	0.7	0.00
Q8K0T7	Protein unc-13 homolog C	*Unc13c*	5	10	0.7	0.01
P68510	14-3-3 protein eta	*Ywhah*	12	152	0.7	0.00
Q9DB72	BTB/POZ domain-containing protein 17	*Btbd17*	8	21	0.7	0.01

P97434	Myosin phosphatase Rho-interacting protein	*Mprip*	4	8	0.7	0.00
Q9ERG2	Striatin-3	*Strn3*	2	5	0.7	0.00
P97445	Voltage-dependent P/Q-type calcium channel subunit alpha-1A	*Cacna1a*	8	23	0.7	0.01
P62075	Mitochondrial import inner membrane translocase subunit Tim13	*Timm13*	3	15	0.7	0.02
Q9EPW0	Type I inositol 3,4-bisphosphate 4-phosphatase	*Inpp4a*	12	41	0.7	0.00
Q8K0T0	Reticulon-1	*Rtn1*	14	63	0.7	0.00
P70302	Stromal interaction molecule 1	*Stim1*	6	11	0.7	0.00
P16283	Anion exchange protein 3	*Slc4a3*	2	6	0.7	0.04
Q4LDD4	Arf-GAP with Rho-GAP domain, ANK repeat and PH domain-containing protein 1	*Arap1*	2	5	0.7	0.00
O55143	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	*Atp2a2*	24	162	0.7	0.00
P16305	Gamma-aminobutyric acid receptor subunit alpha-6	*Gabra6*	2	8	0.7	0.01
Q6PB70	Anoctamin-8	*Ano8*	2	5	0.7	0.01
Q6PA06	Atlastin-2	*Atl2*	7	22	0.7	0.00
P08414	Calcium/calmodulin-dependent protein kinase type IV	*Camk4*	9	27	0.7	0.02
Q5DTL9	Sodium-driven chloride bicarbonate exchanger	*Slc4a10*	12	40	0.6	0.00
Q6WVG3	BTB/POZ domain-containing protein KCTD12	*Kctd12*	9	30	0.6	0.00
P34884	Macrophage migration inhibitory factor	*Mif*	3	15	0.6	0.04
Q3UJU9	Regulator of microtubule dynamics protein 3	*Rmdn3*	7	17	0.6	0.00
O54931	A-kinase anchor protein 2	*Akap2*	2	3	0.6	0.00
B9EJA2	Cortactin-binding protein 2	*Cttnbp2*	5	11	0.6	0.00
P12660	Purkinje cell protein 2	*Pcp2*	3	16	0.6	0.00
Q91YX5	Acyl-CoA:lysophosphatidylglycerol acyltransferase 1	*Lpgat1*	5	21	0.6	0.00
Q69ZT1	Fanconi-associated nuclease 1	*Fan1*	2	2	0.6	0.00
P57759	Endoplasmic reticulum resident protein 29	*Erp29*	9	28	0.6	0.00
P0C605	cGMP-dependent protein kinase 1	*Prkg1*	2	3	0.6	0.00
Q8K1S1	Leucine-rich repeat LGI family member 4	*Lgi4*	4	11	0.6	0.01
O35143	ATPase inhibitor, mitochondrial	*Atpif1*	2	3	0.6	0.01
P62748	Hippocalcin-like protein 1	*Hpcal1*	5	62	0.6	0.00
Q68ED7	CREB-regulated transcription coactivator 1	*Crtc1*	3	7	0.6	0.01
Q8C5W0	Calmin	*Clmn*	2	3	0.6	0.00
Q8R1S4	Metastasis suppressor protein 1	*Mtss1*	3	9	0.6	0.00
Q6ZQ82	Rho GTPase-activating protein 26	*Arhgap26*	9	26	0.6	0.00
P84086	Complexin-2	*Cplx2*	3	33	0.6	0.00
P56564	Excitatory amino acid transporter 1	*Slc1a3*	10	147	0.6	0.00
B2RPU2	Pleckstrin homology domain-containing family D member 1	*Plekhd1*	3	4	0.5	0.00
Q9JHG0	Cerebellin-3	*Cbln3*	2	5	0.5	0.01

P23818	Glutamate receptor 1	*Gria1*	11	39	0.5	0.00
P25911	Tyrosine-protein kinase Lyn	*Lyn*	7	27	0.5	0.00
Q9JM96	Cdc42 effector protein 4	*Cdc42ep4*	6	12	0.5	0.00
Q8BXT1	Regulator of G-protein signaling 8	*Rgs8*	3	6	0.5	0.00
Q80Z38	SH3 and multiple ankyrin repeat domains protein 2	*Shank2*	12	31	0.5	0.00
Q80UP3	Diacylglycerol kinase zeta	*Dgkz*	10	18	0.5	0.00
Q3UH99	Protein shisa-6 homolog	*Shisa6*	6	15	0.5	0.01
Q9ERQ8	Carbonic anhydrase 7	*Ca7*	2	3	0.5	0.00
D3YZU1	SH3 and multiple ankyrin repeat domains protein 1	*Shank1*	25	61	0.5	0.00
Q80YX1	Tenascin	*Tnc*	6	13	0.5	0.00
Q9JJZ2	Tubulin alpha-8 chain	*Tuba8*	4	460	0.5	0.00
Q6WQJ1	Sn1-specific diacylglycerol lipase alpha	*Dagla*	9	20	0.4	0.00
P97772	Metabotropic glutamate receptor 1	*Grm1*	14	35	0.4	0.00
Q8R071	Inositol-trisphosphate 3-kinase A	*Itpka*	4	8	0.4	0.00
O35544	Excitatory amino acid transporter 4	*Slc1a6*	5	69	0.4	0.00
Q3TGF2	Protein FAM107B	*Fam107b*	3	3	0.4	0.00
Q9JMF3	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-13	*Gng13*	4	9	0.4	0.00
Q64518	Sarcoplasmic/endoplasmic reticulum calcium ATPase 3	*Atp2a3*	16	55	0.3	0.00
Q61625	Glutamate receptor ionotropic, delta-2	*Grid2*	10	34	0.3	0.01
Q8BW86	Rho guanine nucleotide exchange factor 33	*Arhgef33*	11	24	0.3	0.00
P63318	Protein kinase C gamma type	*Prkcg*	19	57	0.3	0.00
Q0VEJ0	Centrosomal protein of 76 kDa	*Cep76*	3	3	0.3	0.00
Q0QWG9	Delphilin	*Grid2ip*	5	14	0.3	0.00
P11881	Inositol 1,4,5-trisphosphate receptor type 1	*Itpr1*	73	294	0.3	0.00
Q99JP6	Homer protein homolog 3	*Homer3*	14	45	0.3	0.00
P12658	Calbindin	*Calb1*	15	145	0.2	0.00
P28651	Carbonic anhydrase-related protein	*Ca8*	9	43	0.2	0.00
Q9EQK7	Protein-S-isoprenylcysteine O-methyltransferase	*Icmt*	2	2	0.2	0.00

10.6 Significantly altered proteins in the *woozy* mice skeletal muscles study

UniProt accession	Protein	Gene	Unique Peptides	PSMs	*Woozy*/Wild type	T.TEST
P16110	Galectin-3	*Lgals3*	3	7	3.2	0.02
Q61878	Bone marrow proteoglycan	*Prg2*	2	4	2.7	0.00

P97352	Protein S100-A13	*S100a13*	5	18	2.3	0.04
P50543	Protein S100-A11	*S100a11*	3	4	2.2	0.02
P23927	Alpha-crystallin B chain	*Cryab*	8	86	2.0	0.01
Q9R118	Serine protease HTRA1	*Htra1*	2	3	1.9	0.01
P13595	Neural cell adhesion molecule 1	*Ncam1*	3	5	1.8	0.03
P20029	78 kDa glucose-regulated protein	*Hspa5*	27	235	1.8	0.00
P14602	Heat shock protein beta-1	*Hspb1*	10	84	1.8	0.00
P35385	Heat shock protein beta-7	*Hspb7*	5	73	1.8	0.01
P26350	Prothymosin alpha	*Ptma*	3	10	1.8	0.02
P09541	Myosin light chain 4	*Myl4*	4	276	1.8	0.01
P14069	Protein S100-A6	*S100a6*	3	15	1.7	0.02
Q60854	Serpin B6	*Serpinb6*	18	173	1.7	0.05
Q9CQI6	Coactosin-like protein	*Cotl1*	8	19	1.7	0.00
P29391	Ferritin light chain 1	*Ftl1*	3	22	1.6	0.00
O89053	Coronin-1A	*Coro1a*	8	20	1.6	0.01
Q9CXI5	Mesencephalic astrocyte-derived neurotrophic factor	*Manf*	2	6	1.6	0.01
O35367	Keratocan	*Kera*	9	40	1.6	0.00
Q922F4	Tubulin beta-6 chain	*Tubb6*	3	127	1.6	0.04
P62962	Profilin-1	*Pfn1*	7	67	1.6	0.03
P48036	Annexin A5	*Anxa5*	13	69	1.6	0.01
P10126	Elongation factor 1-alpha 1	*Eef1a1*	5	123	1.6	0.01
Q3U5Q7	UMP-CMP kinase 2, mitochondrial	*Cmpk2*	2	4	1.6	0.01
P50608	Fibromodulin	*Fmod*	8	62	1.6	0.00
P28653	Biglycan	*Bgn*	7	50	1.6	0.01
Q9EQK5	Major vault protein	*Mvp*	18	60	1.5	0.00
P27661	Histone H2AX	*H2afx*	3	67	1.5	0.03
P11352	Glutathione peroxidase 1	*Gpx1*	3	7	1.5	0.02
P17710	Hexokinase-1	*Hk1*	9	40	1.5	0.00
A2AMM0	Muscle-related coiled-coil protein	*Murc*	6	24	1.5	0.00
Q61233	Plastin-2	*Lcp1*	4	15	1.5	0.01
O70373	Xin actin-binding repeat-containing protein 1	*Xirp1*	21	57	1.5	0.00
O08917	Flotillin-1	*Flot1*	3	5	1.5	0.01
P18760	Cofilin-1	*Cfl1*	4	46	1.5	0.01
P08226	Apolipoprotein E	*Apoe*	6	15	1.5	0.04
P15864	Histone H1.2	*Hist1h1c*	3	64	1.5	0.01
Q99KC8	von Willebrand factor A domain-containing protein 5A	*Vwa5a*	13	43	1.5	0.00

P54116	Erythrocyte band 7 integral membrane protein	*Stom*	3	7	1.5	0.00
O09161	Calsequestrin-2	*Casq2*	5	18	1.5	0.03
P10107	Annexin A1	*Anxa1*	8	29	1.5	0.03
Q922R8	Protein disulfide-isomerase A6	*Pdia6*	7	33	1.5	0.00
P97429	Annexin A4	*Anxa4*	13	59	1.5	0.02
Q8R5J9	PRA1 family protein 3	*Arl6ip5*	4	16	1.5	0.03
Q6P5H2	Nestin	*Nes*	8	17	1.5	0.02
P35564	Calnexin	*Canx*	10	37	1.5	0.00
Q3TW96	UDP-N-acetylhexosamine pyrophosphorylase-like protein 1	*Uap1l1*	6	16	1.5	0.01
P43025	Tetranectin	*Clec3b*	3	9	1.5	0.00
P21836	Acetylcholinesterase	*Ache*	2	4	1.5	0.00
Q8C7E7	Starch-binding domain-containing protein 1	*Stbd1*	7	25	1.5	0.01
Q8BMK4	Cytoskeleton-associated protein 4	*Ckap4*	5	7	1.4	0.02
Q8BMD8	Calcium-binding mitochondrial carrier protein SCaMC-1	*Slc25a24*	2	4	1.4	0.02
P09528	Ferritin heavy chain	*Fth1*	5	31	1.4	0.02
P07934	Phosphorylase b kinase gamma catalytic chain, skeletal muscle/heart isoform	*Phkg1*	10	57	0.7	0.01
Q6PD26	GPI transamidase component PIG-S	*Pigs*	2	2	0.7	0.02
Q9JL56	Glycerophosphodiester phosphodiesterase 1	*Gde1*	2	4	0.7	0.00
P70266	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 1	*Pfkfb1*	8	24	0.7	0.05
Q9CXJ4	ATP-binding cassette sub-family B member 8, mitochondrial	*Abcb8*	9	18	0.7	0.01
Q7TSH2	Phosphorylase b kinase regulatory subunit beta	*Phkb*	20	74	0.7	0.01
Q7M729	Sodium channel subunit beta-4	*Scn4b*	3	7	0.7	0.00
Q8C0L9	Glycerophosphocholine phosphodiesterase GPCPD1	*Gpcpd1*	5	10	0.6	0.00
A3KFX0	Cytosolic 5'-nucleotidase 1A	*Nt5c1a*	5	12	0.6	0.00
P56501	Mitochondrial uncoupling protein 3	*Ucp3*	2	5	0.6	0.00
P31154	S-adenosylmethionine decarboxylase proenzyme 1	*Amd1*	3	6	0.4	0.01

10.7 Significantly altered proteins in the index patient skeletal muscle study

UniProt accession	Protein	Gene	Peptides	Spectra	Index/Control	RSD%
P24043	Laminin subunit alpha-2	*LAMA2*	24	34	20.4	2%
P07355	Annexin A2	*ANXA2*	26	61	12.2	14%
P55084	Trifunctional enzyme subunit beta, mitochondrial	*HADHB*	21	43	7.4	5%
P08572	Collagen alpha-2	*COL4A2*	7	13	6.7	18%
P07437	Tubulin beta chain	*TUBB*	22	48	6.2	18%

P06576	ATP synthase subunit beta, mitochondrial	*ATP5B*	31	181	5.8	20%
O75390	Citrate synthase, mitochondrial	*CS*	11	26	5.5	7%
P49753	Acyl-coenzyme A thioesterase 2, mitochondrial	*ACOT2*	8	13	5.2	14%
P07900	Heat shock protein HSP 90-alpha	*HSP90AA1*	7	10	5.2	18%
P62937	Peptidyl-prolyl cis-trans isomerase A	*PPIA*	12	49	4.9	17%
P68366	Tubulin alpha-4A chain	*TUBA4A*	16	46	4.9	10%
O60814	Histone H2B type 1-K	*HIST1H2BK*	13	124	4.9	2%
P12235	ADP/ATP translocase 1	*SLC25A4*	20	47	4.9	18%
P33121	Long-chain-fatty-acid--CoA ligase 1	*ACSL1*	7	8	4.8	2%
P00403	Cytochrome c oxidase subunit 2	*MT-CO2*	4	9	4.6	18%
O00483	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 4	*NDUFA4*	4	5	4.5	2%
P09622	Dihydrolipoyl dehydrogenase, mitochondrial	*DLD*	9	17	4.4	18%
Q9UKS6	Protein kinase C and casein kinase substrate in neurons protein 3	*PACSIN3*	10	18	4.3	8%
P08133	Annexin A6	*ANXA6*	14	30	4.3	7%
Q9P0L0	Vesicle-associated membrane protein-associated protein A	*VAPA*	4	7	4.2	2%
Q9Y277	Voltage-dependent anion-selective channel protein 3	*VDAC3*	9	15	4.2	9%
P22061	Protein-L-isoaspartate	*PCMT1*	5	8	4.1	18%
P63313	Thymosin beta-10	*TMSB10*	3	8	4.1	18%
P06744	Glucose-6-phosphate isomerase	*GPI*	13	41	4.1	6%
P21912	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	*SDHB*	7	17	3.8	16%
Q05639	Elongation factor 1-alpha 2	*EEF1A2*	8	21	3.8	2%
O60662	Kelch-like protein 41	*KLHL41*	26	64	3.8	5%
P48735	Isocitrate dehydrogenase [NADP], mitochondrial	*IDH2*	22	76	3.8	8%
P62805	Histone H4	*HIST1H4A*	13	63	3.7	4%
P11217	Glycogen phosphorylase, muscle form	*PYGM*	69	249	3.6	14%
P21817	Ryanodine receptor 1	*RYR1*	10	14	3.6	18%
A0M8Q6	Ig lambda-7 chain C region	*IGLC7*	6	16	3.6	16%
P68431	Histone H3.1	*HIST1H3A*	12	39	3.6	9%
P07602	Prosaposin	*PSAP*	3	6	3.6	2%
P36957	Dihydrolipoyllysine-residue succinyltransferase, mitochondrial	*DLST*	4	6	3.6	2%
P40926	Malate dehydrogenase, mitochondrial	*MDH2*	21	85	3.6	5%
Q9Y235	Probable C->U-editing enzyme APOBEC-2	*APOBEC2*	8	23	3.5	10%
P04179	Superoxide dismutase [Mn], mitochondrial	*SOD2*	9	17	3.4	14%
P13639	Elongation factor 2	*EEF2*	8	14	3.4	11%
P04908	Histone H2A type 1-B/E	*HIST1H2AB*	11	55	3.3	8%
P17174	Aspartate aminotransferase, cytoplasmic	*GOT1*	15	32	3.3	14%

P28161	Glutathione S-transferase Mu 2	*GSTM2*	4	10	3.3	11%
P15121	Aldose reductase	*AKR1B1*	6	10	3.3	15%
P0C0S8	Histone H2A type 1	*HIST1H2AG*	9	55	3.3	5%
O14983	Sarcoplasmic/endoplasmic reticulum calcium ATPase 1	*ATP2A1*	46	150	3.2	18%
P14618	Pyruvate kinase PKM	*PKM*	45	199	3.2	0%
O75531	Barrier-to-autointegration factor	*BANF1*	7	14	3.2	19%
P30041	Peroxiredoxin-6	*PRDX6*	13	32	3.1	2%
P22695	Cytochrome b-c1 complex subunit 2, mitochondrial	*UQCRC2*	14	22	3.1	2%
P48047	ATP synthase subunit O, mitochondrial	*ATP5O*	13	31	2.9	13%
P62328	Thymosin beta-4	*TMSB4X*	13	35	2.9	15%
Q6ZMU5	Tripartite motif-containing protein 72	*TRIM72*	15	28	2.9	18%
P06733	Alpha-enolase	*ENO1*	27	99	2.8	12%
P54652	Heat shock-related 70 kDa protein 2	*HSPA2*	17	44	2.7	18%
Q9Y623	Myosin-4	*MYH4*	202	1907	2.7	13%
P12882	Myosin-1	*MYH1*	352	3322	2.7	13%
P61088	Ubiquitin-conjugating enzyme E2 N	*UBE2N*	2	3	2.7	2%
Q9UKX3	Myosin-13	*MYH13*	85	750	2.6	1%
P04406	Glyceraldehyde-3-phosphate dehydrogenase	*GAPDH*	55	444	2.6	9%
P09651	Heterogeneous nuclear ribonucleoprotein A1	*HNRNPA1*	4	5	2.6	18%
P16104	Histone H2AX	*H2AFX*	7	38	2.5	3%
P19338	Nucleolin	*NCL*	11	14	2.5	12%
P09104	Gamma-enolase	*ENO2*	14	86	2.5	19%
P0C0S5	Histone H2A.Z	*H2AFZ*	3	18	2.5	13%
P11142	Heat shock cognate 71 kDa protein	*HSPA8*	24	67	2.5	6%
P13073	Cytochrome c oxidase subunit 4 isoform 1, mitochondrial	*COX4I1*	7	15	2.5	15%
Q9NTK5	Obg-like ATPase 1	*OLA1*	3	6	2.4	14%
P0CG47	Polyubiquitin-B	*UBB*	3	10	2.4	8%
P56134	ATP synthase subunit f, mitochondrial	*ATP5J2*	2	4	2.4	2%
O95292	Vesicle-associated membrane protein-associated protein B/C	*VAPB*	3	4	2.4	2%
P61981	14-3-3 protein gamma	*YWHAG*	5	12	2.4	17%
P06732	Creatine kinase M-type	*CKM*	98	854	2.3	19%
P40925	Malate dehydrogenase, cytoplasmic	*MDH1*	18	67	2.3	12%
Q9UKX2	Myosin-2	*MYH2*	335	2762	2.2	18%
P54296	Myomesin-2	*MYOM2*	41	102	2.2	11%
Q06830	Peroxiredoxin-1	*PRDX1*	14	45	2.2	13%
P0DMV8	Heat shock 70 kDa protein 1A	*HSPA1A*	21	48	2.2	16%

P0DMV9	Heat shock 70 kDa protein 1B	*HSPA1B*	21	48	2.2	16%
P06748	Nucleophosmin	*NPM1*	6	12	2.2	2%
P60709	Actin, cytoplasmic 1	*ACTB*	58	1002	2.1	5%
P00558	Phosphoglycerate kinase 1	*PGK1*	32	116	2.1	8%
P10412	Histone H1.4	*HIST1H1E*	16	34	2.1	16%
P63261	Actin, cytoplasmic 2	*ACTG1*	59	985	2.1	5%
P68133	Actin, alpha skeletal muscle	*ACTA1*	127	1899	2.1	3%
Q86TC9	Myopalladin	*MYPN*	3	4	2.1	18%
Q00872	Myosin-binding protein C, slow-type	*MYBPC1*	85	244	2.0	2%
P20929	Nebulin	*NEB*	336	701	2.0	10%
P17661	Desmin	*DES*	47	250	2.0	8%
P23588	Eukaryotic translation initiation factor 4B	*EIF4B*	2	3	0.5	11%
Q00688	Peptidyl-prolyl cis-trans isomerase FKBP3	*FKBP3*	10	15	0.5	13%
P05413	Fatty acid-binding protein, heart	*FABP3*	15	33	0.5	3%
Q9UBY9	Heat shock protein beta-7	*HSPB7*	10	26	0.4	2%
P30049	ATP synthase subunit delta, mitochondrial	*ATP5D*	2	2	0.4	6%
O15273	Telethonin	*TCAP*	3	7	0.2	11%
Q8N3V7	Synaptopodin	*SYNPO*	2	2	0.1	2%

10.8 UPR pathway-associated proteins for the PRM-based targeted assay

UniProt Accession	Protein	Gene	Peptide Sequence	Precursor m/z
P11021	78 kDa glucose-regulated protein	*GRP78*	VEIIANDQGNR	614.8177
P11021	78 kDa glucose-regulated protein	*GRP78*	TWNDPSVQQDIK	715.8492
P11021	78 kDa glucose-regulated protein	*GRP78*	DNHLLGTFDLTGIPPAPR	645.3425
P11021	78 kDa glucose-regulated protein	*GRP78*	LTPEEIER	493.7613
P11021	78 kDa glucose-regulated protein	*GRP78*	IEWLESHQDADIEDFK	658.9742
P14625	Endoplasmin	*ENPL*	ELISNASDALDK	638.3250
P14625	Endoplasmin	*ENPL*	SILFVPTSAPR	594.3428
P14625	Endoplasmin	*ENPL*	GLFDEYGSK	508.2402
P14625	Endoplasmin	*ENPL*	LGVIEDHSNR	380.5316
P14625	Endoplasmin	*ENPL*	LSLNIDPDAK	543.2955
Q9Y4L1	Hypoxia up-regulated protein 1	*HYOU1*	FPEHELTFDPQR	505.9124
Q9Y4L1	Hypoxia up-regulated protein 1	*HYOU1*	LAGLFNEQR	524.2827

Q9Y4L1	Hypoxia up-regulated protein 1	*HYOU1*	DAVVYPILVEFTR	761.4192
Q9Y4L1	Hypoxia up-regulated protein 1	*HYOU1*	EVEEEPGIHSLK	456.2331
Q9Y4L1	Hypoxia up-regulated protein 1	*HYOU1*	AEAGPEGVAPAPEGEK	754.8650
O94979	Protein transport protein Sec31A	*SC31A*	TQPPEDISC[+57]IAWNR	843.8989
O94979	Protein transport protein Sec31A	*SC31A*	NPAVLSAASFDGR	652.8333
O94979	Protein transport protein Sec31A	*SC31A*	LVTFENVR	489.2744
O94979	Protein transport protein Sec31A	*SC31A*	AQDGSHPLSLQDLIEK	584.3039
P02545	Prelamin-A/C	*LMNA*	EGDLIAAQAR	522.2776
P02545	Prelamin-A/C	*LMNA*	EAALSTALSEK	560.2982
P02545	Prelamin-A/C	*LMNA*	TLEGELHDLR	591.8093
P02545	Prelamin-A/C	*LMNA*	LADALQELR	514.7904
P05198	Eukaryotic translation initiation factor 2 subunit 1	*IF2A*	VSPEEAIK	436.7398
P05198	Eukaryotic translation initiation factor 2 subunit 1	*IF2A*	ADIEVAC[+57]YGYEGIDAVK	936.9378
P05198	Eukaryotic translation initiation factor 2 subunit 1	*IF2A*	INLIAPPR	447.2820
P05198	Eukaryotic translation initiation factor 2 subunit 1	*IF2A*	VVTDTDETELAR	674.8332
P27824	Calnexin	*CALX*	APVPTGEVYFADSFDR	885.9203
P27824	Calnexin	*CALX*	GTLSGWILSK	531.3031
P27824	Calnexin	*CALX*	TPELNLDQFHDK	728.8570
P27824	Calnexin	*CALX*	AEEDEILNR	544.7646
P38646	Stress-70 protein, mitochondrial	*GRP75*	DAGQISGLNVLR	621.8437
P38646	Stress-70 protein, mitochondrial	*GRP75*	VINEPTAAALAYGLDK	823.4434
P38646	Stress-70 protein, mitochondrial	*GRP75*	AQFEGIVTDLIR	681.3748
P38646	Stress-70 protein, mitochondrial	*GRP75*	LLGQFTLIGIPPAPR	796.9798
P55072	Transitional endoplasmic reticulum ATPase	*TERA*	EVDIGIPDATGR	621.8199
P55072	Transitional endoplasmic reticulum ATPase	*TERA*	LEILQIHTK	547.8320
P55072	Transitional endoplasmic reticulum ATPase	*TERA*	WALSQSNPSALR	665.3491
P55072	Transitional endoplasmic reticulum ATPase	*TERA*	DVDLEFLAK	525.2793
P04406	Glyceraldehyde-3-phosphate dehydrogenase	*G3P*	IISNASC[+57]TTNC[+57]LAPLAK	611.9781
P04406	Glyceraldehyde-3-phosphate dehydrogenase	*G3P*	VPTANVSVVDLTC[+57]R	510.9363
P04406	Glyceraldehyde-3-phosphate dehydrogenase	*G3P*	LISWYDNEFGYSNR	588.6056
P27797	Calreticulin	*CALR*	FYALSASFEPFSNK	804.3907
P27797	Calreticulin	*CALR*	GQTLVVQFTVK	610.3559
P27797	Calreticulin	*CALR*	HEQNIDC[+57]GGGYVK	738.8304
Q14554	Protein disulfide-isomerase A5	*PDIA5*	SEVAAENHLR	563.2860
Q14554	Protein disulfide-isomerase A5	*PDIA5*	GPPLWEEDPGAK	648.3170
Q14554	Protein disulfide-isomerase A5	*PDIA5*	GFPTIC[+57]YFEK	631.2997

O75190	DnaJ homolog subfamily B member 6	*DNJB6*	VEVEEDGQLK	573.2879
O75190	DnaJ homolog subfamily B member 6	*DNJB6*	APGPWDPLASAAGLK	725.8881
P04792	Heat shock protein beta-1	*HB1*	LFDQAFGLPR	388.5451
P04792	Heat shock protein beta-1	*HB1*	LATQSNEITIPVTFESR	636.0021
P07237	Protein disulfide-isomerase	*PDIA1*	ALAPEYAK	431.7371
P07237	Protein disulfide-isomerase	*PDIA1*	VDATEESDLAQQYGVR	890.9210
P49257	Protein ERGIC-53	*LMAN1*	GHPDLQGQPAEEIFESVGDR	1091.0140
P49257	Protein ERGIC-53	*LMAN1*	DIDNLVQR	486.7591
P61619	Protein transport protein Sec61 subunit alpha isoform 1	*S61A1*	IIEVGDTPK	486.2740
P61619	Protein transport protein Sec61 subunit alpha isoform 1	*S61A1*	AFSPTTVNTGR	575.7962
Q13217	DnaJ homolog subfamily C member 3	*DNJC3*	SQALNAFGSGDYTAAIAFLDK	720.6918
Q13217	DnaJ homolog subfamily C member 3	*DNJC3*	LIESAEELIR	586.8297
Q99442	Translocation protein SEC62	*SEC62*	VDYFIASK	471.7502
Q99442	Translocation protein SEC62	*SEC62*	AVDC[+57]LLDSK	510.7551
Q99497	Protein DJ-1	*PARK7*	DGLILTSR	437.7533
Q99497	Protein DJ-1	*PARK7*	GPGTSFEFALAIVEALNGK	641.0070
Q9H173	Nucleotide exchange factor SIL1	*SIL1*	EFALTNPEK	524.7691
Q9H173	Nucleotide exchange factor SIL1	*SIL1*	LGGLQVLR	428.2742
Q9NYU2	UDP-glucose:glycoprotein glucosyltransferase 1	*UGGG1*	LNIQPSEADYAVDIR	852.4336
Q9NYU2	UDP-glucose:glycoprotein glucosyltransferase 1	*UGGG1*	IEYQFFEDR	623.7906
Q9UBS4	DnaJ homolog subfamily B member 11	*DJB11*	FQDLGAAYEVLSDSEK	886.4229
Q9UBS4	DnaJ homolog subfamily B member 11	*DJB11*	TLEVEIEPGVR	621.3404
O75460	Serine/threonine-protein kinase/endoribonuclease IRE1	*ERN1*	LPFTIPELVQASPC[+57]R	864.4611
P08240	Signal recognition particle receptor subunit alpha	*SRPR*	NQGFDVVLVDTAGR	745.8835
Q96RQ1	Endoplasmic reticulum-Golgi intermediate compartment protein 2	*ERGI2*	IDHLSFGELVPAIINPLDGTEK	793.4249
Q9Y282	Endoplasmic reticulum-Golgi intermediate compartment protein 3	ERGI3	VEVTVFDPDSLDPDR	852.4098

10.9 Significantly altered proteins in the SIL1-depleted HEK293 cell line

UniProt Accession	Protein	Gene	ΔSIL1/scr
Q96FW1	Ubiquitin thioesterase OTUB1	*OTUB1*	4.9
Q01581	Hydroxymethylglutaryl-CoA synthase, cytoplasmic	*HMGCS1*	4.6
Q9NY33	Dipeptidyl peptidase 3	*DPP3*	4.6
Q96C86	m7GpppX diphosphatase	*DCPS*	4.6

Q9H2J4	Phosducin-like protein 3	*PDCL3*	4.5
P04818	Thymidylate synthase	*TYMS*	4.2
Q96C90	Protein phosphatase 1 regulatory subunit 14B	*PPP1R14B*	4.2
P02765	Alpha-2-HS-glycoprotein	*AHSG*	4.1
Q08623	Pseudouridine-5'-monophosphatase	*HDHD1*	4.0
P08397	Porphobilinogen deaminase	*HMBS*	3.9
Q15102	Platelet-activating factor acetylhydrolase IB subunit gamma	*PAFAH1B3*	3.6
Q99757	Thioredoxin, mitochondrial	*TXN2*	3.4
O60701	UDP-glucose 6-dehydrogenase	*UGDH*	3.4
P00374	Dihydrofolate reductase	*DHFR*	3.4
P19623	Spermidine synthase	*SRM*	3.3
O15294	UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit	*OGT*	3.2
Q9UHY7	Enolase-phosphatase E1	*ENOPH1*	3.1
P07339	Cathepsin D	*CTSD*	3.1
P00338	L-lactate dehydrogenase A chain	*LDHA*	3.1
Q9Y570	Protein phosphatase methylesterase 1	*PPME1*	3.1
P29218	Inositol monophosphatase 1	*IMPA1*	3.1
O75874	Isocitrate dehydrogenase [NADP] cytoplasmic	*IDH1*	3.1
P68402	Platelet-activating factor acetylhydrolase IB subunit beta	*PAFAH1B2*	3.0
Q4G0N4	NAD kinase 2, mitochondrial	*NADK2*	3.0
Q9UHD1	Cysteine and histidine-rich domain-containing protein 1	*CHORDC1*	3.0
P31350	Ribonucleoside-diphosphate reductase subunit M2	*RRM2*	2.9
P15121	Aldose reductase	*AKR1B1*	2.9
Q9UNZ2	NSFL1 cofactor p47	*NSFL1C*	2.9
O00762	Ubiquitin-conjugating enzyme E2 C	*UBE2C*	2.9
P20839	Inosine-5'-monophosphate dehydrogenase 1	*IMPDH1*	2.9
Q96G03	Phosphoglucomutase-2	*PGM2*	2.9
P30044	Peroxiredoxin-5, mitochondrial	*PRDX5*	2.9
Q4VC31	Coiled-coil domain-containing protein 58	*CCDC58*	2.8
Q9Y5J9	Mitochondrial import inner membrane translocase subunit Tim8 B	*TIMM8B*	2.8
Q96GD0	Pyridoxal phosphate phosphatase	*PDXP*	2.8
P52758	Ribonuclease UK114	*HRSP12*	2.8
O00193	Small acidic protein	*SMAP*	2.8
P61916	Epididymal secretory protein E1	*NPC2*	2.8
Q5U5X0	Complex III assembly factor LYRM7	*LYRM7*	2.8
P54105	Methylosome subunit pICln	*CLNS1A*	2.7

O75391	Sperm-associated antigen 7	*SPAG7*	2.7
Q9NRX4	14 kDa phosphohistidine phosphatase	*PHPT1*	2.7
P07741	Adenine phosphoribosyltransferase	*APRT*	2.7
Q7Z4W1	L-xylulose reductase	*DCXR*	2.7
Q15257	Serine/threonine-protein phosphatase 2A activator	*PPP2R4*	2.7
Q96RP9	Elongation factor G, mitochondrial	*GFM1*	2.7
P25787	Proteasome subunit alpha type-2	*PSMA2*	2.6
P60174	Triosephosphate isomerase	*TPI1*	2.6
Q9UL25	Ras-related protein Rab-21	*RAB21*	2.6
Q9HC38	Glyoxalase domain-containing protein 4	*GLOD4*	2.6
Q9BY32	Inosine triphosphate pyrophosphatase	*ITPA*	2.6
Q16543	Hsp90 co-chaperone Cdc37	*CDC37*	2.6
Q9NUQ9	Protein FAM49B	*FAM49B*	2.6
P49773	Histidine triad nucleotide-binding protein 1	*HINT1*	2.6
P13797	Plastin-3	*PLS3*	2.6
P28072	Proteasome subunit beta type-6	*PSMB6*	2.6
P27144	Adenylate kinase 4, mitochondrial	*AK4*	2.6
P18669	Phosphoglycerate mutase 1	*PGAM1*	2.6
O95336	6-phosphogluconolactonase	*PGLS*	2.6
Q9NVS9	Pyridoxine-5'-phosphate oxidase	*PNPO*	2.6
P52209	6-phosphogluconate dehydrogenase, decarboxylating	*PGD*	2.6
O43768	Alpha-endosulfine	*ENSA*	2.6
O94925	Glutaminase kidney isoform, mitochondrial	*GLS*	2.6
Q5TDH0	Protein DDI1 homolog 2	*DDI2*	2.5
P46108	Adapter molecule crk	*CRK*	2.5
P31939	Bifunctional purine biosynthesis protein PURH	*ATIC*	2.5
Q9HB07	UPF0160 protein MYG1, mitochondrial	*C12orf10*	2.5
P07195	L-lactate dehydrogenase B chain	*LDHB*	2.5
P49720	Proteasome subunit beta type-3	*PSMB3*	2.5
Q9BWD1	Acetyl-CoA acetyltransferase, cytosolic	*ACAT2*	2.5
P17174	Aspartate aminotransferase, cytoplasmic	*GOT1*	2.5
P09211	Glutathione S-transferase P	*GSTP1*	2.5
Q99436	Proteasome subunit beta type-7	*PSMB7*	2.5
Q96KP4	Cytosolic non-specific dipeptidase	*CNDP2*	2.5
Q9UBQ7	Glyoxylate reductase/hydroxypyruvate reductase	*GRHPR*	2.5
P06733	Alpha-enolase	*ENO1*	2.5

P06132	Uroporphyrinogen decarboxylase	*UROD*	2.5
P51003	Poly	*PAPOLA*	2.5
Q13492	Phosphatidylinositol-binding clathrin assembly protein	*PICALM*	2.5
P30046	D-dopachrome decarboxylase	*DDT*	2.5
Q9NRN7	L-aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase	*AASDHPPT*	2.5
P37837	Transaldolase	*TALDO1*	2.5
P62136	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	*PPP1CA*	2.5
Q99497	Protein DJ-1	*PARK7*	2.5
Q5TFE4	5'-nucleotidase domain-containing protein 1	*NT5DC1*	2.5
P23921	Ribonucleoside-diphosphate reductase large subunit	*RRM1*	2.4
P16152	Carbonyl reductase [NADPH] 1	*CBR1*	2.4
P15374	Ubiquitin carboxyl-terminal hydrolase isozyme L3	*UCHL3*	2.4
Q9NZL9	Methionine adenosyltransferase 2 subunit beta	*MAT2B*	2.4
P25789	Proteasome subunit alpha type-4	*PSMA4*	2.4
Q9BTT0	Acidic leucine-rich nuclear phosphoprotein 32 family member E	*ANP32E*	2.4
P24752	Acetyl-CoA acetyltransferase, mitochondrial	*ACAT1*	2.4
P49189	4-trimethylaminobutyraldehyde dehydrogenase	*ALDH9A1*	2.4
Q7KZ85	Transcription elongation factor SPT6	*SUPT6H*	2.4
O43598	2'-deoxynucleoside 5'-phosphate N-hydrolase 1	*DNPH1*	2.4
O15305	Phosphomannomutase 2	*PMM2*	2.4
P48163	NADP-dependent malic enzyme	*ME1*	2.4
O76003	Glutaredoxin-3	*GLRX3*	2.4
P12277	Creatine kinase B-type	*CKB*	2.4
P31153	S-adenosylmethionine synthase isoform type-2	*MAT2A*	2.4
P06865	Beta-hexosaminidase subunit alpha	*HEXA*	2.4
Q01518	Adenylyl cyclase-associated protein 1	*CAP1*	2.4
P04075	Fructose-bisphosphate aldolase A	*ALDOA*	2.4
Q01105	Protein SET	*SET*	2.4
Q5TBB1	Ribonuclease H2 subunit B	*RNASEH2B*	2.4
P10768	S-formylglutathione hydrolase	*ESD*	2.4
Q8N7H5	RNA polymerase II-associated factor 1 homolog	*PAF1*	2.4
P21283	V-type proton ATPase subunit C 1	*ATP6V1C1*	2.4
P60983	Glia maturation factor beta	*GMFB*	2.4
Q9H910	Hematological and neurological expressed 1-like protein	*HN1L*	2.4
P18206	Vinculin	*VCL*	2.3
P09972	Fructose-bisphosphate aldolase C	*ALDOC*	2.3

Q9NR45	Sialic acid synthase	*NANS*	2.3
O75347	Tubulin-specific chaperone A	*TBCA*	2.3
O00154	Cytosolic acyl coenzyme A thioester hydrolase	*ACOT7*	2.3
Q8TDP1	Ribonuclease H2 subunit C	*RNASEH2C*	2.3
P29401	Transketolase	*TKT*	2.3
P31948	Stress-induced-phosphoprotein 1	*STIP1*	2.3
Q9Y2Z0	Suppressor of G2 allele of SKP1 homolog	*SUGT1*	2.3
P40926	Malate dehydrogenase, mitochondrial	*MDH2*	2.3
P04406	Glyceraldehyde-3-phosphate dehydrogenase	*GAPDH*	2.3
Q13526	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1	*PIN1*	2.3
Q9P287	BRCA2 and CDKN1A-interacting protein	*BCCIP*	2.3
Q92688	Acidic leucine-rich nuclear phosphoprotein 32 family member B	*ANP32B*	2.3
P25786	Proteasome subunit alpha type-1	*PSMA1*	2.3
P23526	Adenosylhomocysteinase	*AHCY*	2.3
Q15185	Prostaglandin E synthase 3	*PTGES3*	2.3
P30086	Phosphatidylethanolamine-binding protein 1	*PEBP1*	2.3
P62942	Peptidyl-prolyl cis-trans isomerase FKBP1A	*FKBP1A*	2.3
P06744	Glucose-6-phosphate isomerase	*GPI*	2.3
P58546	Myotrophin	*MTPN*	2.3
O75792	Ribonuclease H2 subunit A	*RNASEH2A*	2.3
Q13126	S-methyl-5'-thioadenosine phosphorylase	*MTAP*	2.3
O43488	Aflatoxin B1 aldehyde reductase member 2	*AKR7A2*	2.3
Q92530	Proteasome inhibitor PI31 subunit	*PSMF1*	2.3
Q92820	Gamma-glutamyl hydrolase	*GGH*	2.3
P55263	Adenosine kinase	*ADK*	2.3
Q9BTE7	DCN1-like protein 5	*DCUN1D5*	2.3
O75608	Acyl-protein thioesterase 1	*LYPLA1*	2.3
Q04760	Lactoylglutathione lyase	*GLO1*	2.2
O00584	Ribonuclease T2	*RNASET2*	2.2
Q7Z4H3	HD domain-containing protein 2	*HDDC2*	2.2
P60900	Proteasome subunit alpha type-6	*PSMA6*	2.2
P22392	Nucleoside diphosphate kinase B	*NME2*	2.2
P12955	Xaa-Pro dipeptidase	*PEPD*	2.2
P78417	Glutathione S-transferase omega-1	*GSTO1*	2.2
P14618	Pyruvate kinase PKM	*PKM*	2.2
P40925	Malate dehydrogenase, cytoplasmic	*MDH1*	2.2

Q13867	Bleomycin hydrolase	*BLMH*	2.2
P28074	Proteasome subunit beta type-5	*PSMB5*	2.2
P14174	Macrophage migration inhibitory factor	*MIF*	2.2
Q14166	Tubulin--tyrosine ligase-like protein 12	*TTLL12*	2.2
P41240	Tyrosine-protein kinase CSK	*CSK*	2.2
P22314	Ubiquitin-like modifier-activating enzyme 1	*UBA1*	2.2
P24666	Low molecular weight phosphotyrosine protein phosphatase	*ACP1*	2.2
O95861	3'(2'),5'-bisphosphate nucleotidase 1	*BPNT1*	2.2
P00491	Purine nucleoside phosphorylase	*PNP*	2.2
P14324	Farnesyl pyrophosphate synthase	*FDPS*	2.2
P54725	UV excision repair protein RAD23 homolog A	*RAD23A*	2.2
Q01469	Fatty acid-binding protein, epidermal	*FABP5*	2.2
O00303	Eukaryotic translation initiation factor 3 subunit F	*EIF3F*	2.2
Q7Z5L9	Interferon regulatory factor 2-binding protein 2	*IRF2BP2*	2.2
Q9NQR4	Omega-amidase NIT2	*NIT2*	2.2
P61604	10 kDa heat shock protein, mitochondrial	*HSPE1*	2.2
Q9H993	UPF0364 protein C6orf211	*C6orf211*	2.2
P28838	Cytosol aminopeptidase	*LAP3*	2.2
Q16836	Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	*HADH*	2.2
Q96CN7	Isochorismatase domain-containing protein 1	*ISOC1*	2.2
O94903	Proline synthase co-transcribed bacterial homolog protein	*PROSC*	2.2
Q9NZL4	Hsp70-binding protein 1	*HSPBP1*	2.2
Q12765	Secernin-1	*SCRN1*	2.2
P42771	Cyclin-dependent kinase inhibitor 2A, isoforms 1/2/3	*CDKN2A*	2.2
P63104	14-3-3 protein zeta/delta	*YWHAZ*	2.1
P39687	Acidic leucine-rich nuclear phosphoprotein 32 family member A	*ANP32A*	2.1
O00148	ATP-dependent RNA helicase DDX39A	*DDX39A*	2.1
P28070	Proteasome subunit beta type-4	*PSMB4*	2.1
P54819	Adenylate kinase 2, mitochondrial	*AK2*	2.1
O14737	Programmed cell death protein 5	*PDCD5*	2.1
P29692	Elongation factor 1-delta	*EEF1D*	2.1
O00170	AH receptor-interacting protein	*AIP*	2.1
P55786	Puromycin-sensitive aminopeptidase	*NPEPPS*	2.1
O14745	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	*SLC9A3R1*	2.1
O43719	HIV Tat-specific factor 1	*HTATSF1*	2.1
P00505	Aspartate aminotransferase, mitochondrial	*GOT2*	2.1

O00299	Chloride intracellular channel protein 1	*CLIC1*	2.1
P28066	Proteasome subunit alpha type-5	*PSMA5*	2.1
O60493	Sorting nexin-3	*SNX3*	2.1
P08243	Asparagine synthetase [glutamine-hydrolyzing]	*ASNS*	2.1
Q5JRX3	Presequence protease, mitochondrial	*PITRM1*	2.1
Q99613	Eukaryotic translation initiation factor 3 subunit C	*EIF3C*	2.1
P43034	Platelet-activating factor acetylhydrolase IB subunit alpha	*PAFAH1B1*	2.1
P07737	Profilin-1	*PFN1*	2.1
Q9HAV7	GrpE protein homolog 1, mitochondrial	*GRPEL1*	2.1
P04181	Ornithine aminotransferase, mitochondrial	*OAT*	2.1
P48147	Prolyl endopeptidase	*PREP*	2.1
O00273	DNA fragmentation factor subunit alpha	*DFFA*	2.1
P30520	Adenylosuccinate synthetase isozyme 2	*ADSS*	2.1
P08238	Heat shock protein HSP 90-beta	*HSP90AB1*	2.1
Q15813	Tubulin-specific chaperone E	*TBCE*	2.1
P00558	Phosphoglycerate kinase 1	*PGK1*	2.1
O60763	General vesicular transport factor p115	*USO1*	2.1
Q16763	Ubiquitin-conjugating enzyme E2 S	*UBE2S*	2.1
Q13347	Eukaryotic translation initiation factor 3 subunit I	*EIF3I*	2.1
Q8N8S7	Protein enabled homolog	*ENAH*	2.1
P12004	Proliferating cell nuclear antigen	*PCNA*	2.1
P61970	Nuclear transport factor 2	*NUTF2*	2.0
P48637	Glutathione synthetase	*GSS*	2.0
Q12972	Nuclear inhibitor of protein phosphatase 1	*PPP1R8*	2.0
P61106	Ras-related protein Rab-14	*RAB14*	2.0
P07954	Fumarate hydratase, mitochondrial	*FH*	2.0
P52788	Spermine synthase	*SMS*	2.0
P61086	Ubiquitin-conjugating enzyme E2 K	*UBE2K*	2.0
Q13564	NEDD8-activating enzyme E1 regulatory subunit	*NAE1*	2.0
O60888	Protein CutA	*CUTA*	2.0
Q14240	Eukaryotic initiation factor 4A-II	*EIF4A2*	2.0
O43865	Putative adenosylhomocysteinase 2	*AHCYL1*	2.0
Q92598	Heat shock protein 105 kDa	*HSPH1*	2.0
P49419	Alpha-aminoadipic semialdehyde dehydrogenase	*ALDH7A1*	2.0
Q9UJU6	Drebrin-like protein	*DBNL*	2.0
O15067	Phosphoribosylformylglycinamidine synthase	*PFAS*	2.0

P49321	Nuclear autoantigenic sperm protein	*NASP*	2.0
P49354	Protein farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha	*FNTA*	2.0
P34932	Heat shock 70 kDa protein 4	*HSPA4*	2.0
P45974	Ubiquitin carboxyl-terminal hydrolase 5	*USP5*	2.0
Q13177	Serine/threonine-protein kinase PAK 2	*PAK2*	2.0
P54727	UV excision repair protein RAD23 homolog B	*RAD23B*	2.0
P23381	Tryptophan--tRNA ligase, cytoplasmic	*WARS*	2.0
P26639	Threonine--tRNA ligase, cytoplasmic	*TARS*	2.0
P11766	Alcohol dehydrogenase class-3	*ADH5*	2.0
O14818	Proteasome subunit alpha type-7	*PSMA7*	2.0
P26641	Elongation factor 1-gamma	*EEF1G*	2.0
P09960	Leukotriene A-4 hydrolase	*LTA4H*	2.0
Q96PZ0	Pseudouridylate synthase 7 homolog	*PUS7*	2.0
P54578	Ubiquitin carboxyl-terminal hydrolase 14	*USP14*	2.0
Q08257	Quinone oxidoreductase	*CRYZ*	2.0
Q9Y262	Eukaryotic translation initiation factor 3 subunit L	*EIF3L*	2.0
Q13404	Ubiquitin-conjugating enzyme E2 variant 1	*UBE2V1*	2.0
P50395	Rab GDP dissociation inhibitor beta	*GDI2*	1.9
Q9NVA2	Septin-11	*Sep 11*	1.9
P26440	Isovaleryl-CoA dehydrogenase, mitochondrial	*IVD*	1.9
P68036	Ubiquitin-conjugating enzyme E2 L3	*UBE2L3*	1.9
P35241	Radixin	*RDX*	1.9
Q99426	Tubulin-folding cofactor B	*TBCB*	1.9
O75153	Clustered mitochondria protein homolog	*CLUH*	1.9
Q02790	Peptidyl-prolyl cis-trans isomerase FKBP4	*FKBP4*	1.9
Q96EK6	Glucosamine 6-phosphate N-acetyltransferase	*GNPNAT1*	1.9
Q9H4A4	Aminopeptidase B	*RNPEP*	1.9
P18615	Negative elongation factor E	*NELFE*	1.9
P23528	Cofilin-1	*CFL1*	1.9
P15531	Nucleoside diphosphate kinase A	*NME1*	1.9
Q9HC35	Echinoderm microtubule-associated protein-like 4	*EML4*	1.9
O15498	Synaptobrevin homolog YKT6	*YKT6*	1.9
Q9H773	dCTP pyrophosphatase 1	*DCTPP1*	1.9
Q7L2H7	Eukaryotic translation initiation factor 3 subunit M	*EIF3M*	1.9
P49366	Deoxyhypusine synthase	*DHPS*	1.9
P49915	GMP synthase [glutamine-hydrolyzing]	*GMPS*	1.9

Q9NR33	DNA polymerase epsilon subunit 4	*POLE4*	1.9
Q9GZT8	Putative GTP cyclohydrolase 1 type 2 NIF3L1	*NIF3L1*	1.9
Q99733	Nucleosome assembly protein 1-like 4	*NAP1L4*	1.9
Q15008	26S proteasome non-ATPase regulatory subunit 6	*PSMD6*	1.9
O43765	Small glutamine-rich tetratricopeptide repeat-containing protein alpha	*SGTA*	1.9
O75821	Eukaryotic translation initiation factor 3 subunit G	*EIF3G*	1.9
P20618	Proteasome subunit beta type-1	*PSMB1*	1.9
P32119	Peroxiredoxin-2	*PRDX2*	1.9
Q5VW32	BRO1 domain-containing protein BROX	*BROX*	1.9
Q96CX2	BTB/POZ domain-containing protein KCTD12	*KCTD12*	1.9
P31946	14-3-3 protein beta/alpha	*YWHAB*	1.9
P30084	Enoyl-CoA hydratase, mitochondrial	*ECHS1*	1.9
P33316	Deoxyuridine 5'-triphosphate nucleotidohydrolase, mitochondrial	*DUT*	1.9
P23919	Thymidylate kinase	*DTYMK*	1.8
Q9Y490	Talin-1	*TLN1*	1.8
P00492	Hypoxanthine-guanine phosphoribosyltransferase	*HPRT1*	1.8
O14929	Histone acetyltransferase type B catalytic subunit	*HAT1*	1.8
Q15181	Inorganic pyrophosphatase	*PPA1*	1.8
P60842	Eukaryotic initiation factor 4A-I	*EIF4A1*	1.8
Q16851	UTP--glucose-1-phosphate uridylyltransferase	*UGP2*	1.8
Q9NP81	Serine--tRNA ligase, mitochondrial	*SARS2*	1.8
O00487	26S proteasome non-ATPase regulatory subunit 14	*PSMD14*	1.8
Q9Y617	Phosphoserine aminotransferase	*PSAT1*	1.8
Q9NTK5	Obg-like ATPase 1	*OLA1*	1.8
Q9Y266	Nuclear migration protein nudC	*NUDC*	1.8
Q99961	Endophilin-A2	*SH3GL1*	1.8
P60981	Destrin	*DSTN*	1.8
O15372	Eukaryotic translation initiation factor 3 subunit H	*EIF3H*	1.8
P23193	Transcription elongation factor A protein 1	*TCEA1*	1.8
P62937	Peptidyl-prolyl cis-trans isomerase A	*PPIA*	1.8
O15212	Prefoldin subunit 6	*PFDN6*	1.8
Q8TBC4	NEDD8-activating enzyme E1 catalytic subunit	*UBA3*	1.8
Q14232	Translation initiation factor eIF-2B subunit alpha	*EIF2B1*	1.8
Q14152	Eukaryotic translation initiation factor 3 subunit A	*EIF3A*	1.7
Q2TAL8	Glutamine-rich protein 1	*QRICH1*	1.7
Q9Y5L4	Mitochondrial import inner membrane translocase subunit Tim13	*TIMM13*	1.7

P51665	26S proteasome non-ATPase regulatory subunit 7	*PSMD7*	1.7
P61758	Prefoldin subunit 3	*VBP1*	1.7
P43487	Ran-specific GTPase-activating protein	*RANBP1*	1.7
Q9C0C9	E2/E3 hybrid ubiquitin-protein ligase UBE2O	*UBE2O*	1.7
P62258	14-3-3 protein epsilon	*YWHAE*	1.7
P67870	Casein kinase II subunit beta	*CSNK2B*	1.7
P11172	Uridine 5'-monophosphate synthase	*UMPS*	1.7
P62993	Growth factor receptor-bound protein 2	*GRB2*	1.7
Q15435	Protein phosphatase 1 regulatory subunit 7	*PPP1R7*	1.7
Q9Y520	Protein PRRC2C	*PRRC2C*	1.7
P61160	Actin-related protein 2	*ACTR2*	1.7
Q9GZZ9	Ubiquitin-like modifier-activating enzyme 5	*UBA5*	1.7
P50570	Dynamin-2	*DNM2*	1.7
P11802	Cyclin-dependent kinase 4	*CDK4*	1.7
P49327	Fatty acid synthase	*FASN*	1.7
Q15691	Microtubule-associated protein RP/EB family member 1	*MAPRE1*	1.7
P55036	26S proteasome non-ATPase regulatory subunit 4	*PSMD4*	1.6
P12270	Nucleoprotein TPR	*TPR*	1.6
Q5VT52	Regulation of nuclear pre-mRNA domain-containing protein 2	*RPRD2*	1.6
Q96AC1	Fermitin family homolog 2	*FERMT2*	1.6
O43242	26S proteasome non-ATPase regulatory subunit 3	*PSMD3*	1.6
P38646	Stress-70 protein, mitochondrial	*HSPA9*	1.6
Q9P289	Serine/threonine-protein kinase MST4	*MST4*	1.6
P43686	26S protease regulatory subunit 6B	*PSMC4*	1.6
Q13619	Cullin-4A	*CUL4A*	1.6
Q6P1J9	Parafibromin	*CDC73*	1.6
P09012	U1 small nuclear ribonucleoprotein A	*SNRPA*	0.6
Q8IXT5	RNA-binding protein 12B	*RBM12B*	0.6
P51570	Galactokinase	*GALK1*	0.6
Q9NZI8	Insulin-like growth factor 2 mRNA-binding protein 1	*IGF2BP1*	0.6
P50402	Emerin	*EMD*	0.6
P53582	Methionine aminopeptidase 1	*METAP1*	0.6
P04843	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	*RPN1*	0.6
P61026	Ras-related protein Rab-10	*RAB10*	0.6
P51571	Translocon-associated protein subunit delta	*SSR4*	0.6
Q8NI36	WD repeat-containing protein 36	*WDR36*	0.6

Q96CW5	Gamma-tubulin complex component 3	*TUBGCP3*	0.6
Q92616	Translational activator GCN1	*GCN1L1*	0.6
P35613	Basigin	*BSG*	0.6
Q14165	Malectin	*MLEC*	0.6
Q8IWA0	WD repeat-containing protein 75	*WDR75*	0.6
O43795	Unconventional myosin-Ib	*MYO1B*	0.6
O96008	Mitochondrial import receptor subunit TOM40 homolog	*TOMM40*	0.6
Q9Y679	Ancient ubiquitous protein 1	*AUP1*	0.6
P49207	60S ribosomal protein L34	*RPL34*	0.6
O00264	Membrane-associated progesterone receptor component 1	*PGRMC1*	0.6
P48047	ATP synthase subunit O, mitochondrial	*ATP5O*	0.6
Q00577	Transcriptional activator protein Pur-alpha	*PURA*	0.6
O95292	Vesicle-associated membrane protein-associated protein B/C	*VAPB*	0.6
Q96QD9	UAP56-interacting factor	*FYTTD1*	0.6
Q9Y512	Sorting and assembly machinery component 50 homolog	*SAMM50*	0.6
P62995	Transformer-2 protein homolog beta	*TRA2B*	0.6
Q9Y3B3	Transmembrane emp24 domain-containing protein 7	*TMED7*	0.6
Q8WTT2	Nucleolar complex protein 3 homolog	*NOC3L*	0.6
P07910	Heterogeneous nuclear ribonucleoproteins C1/C2	*HNRNPC*	0.6
Q10471	Polypeptide N-acetylgalactosaminyltransferase 2	*GALNT2*	0.6
Q16891	Mitochondrial inner membrane protein	*IMMT*	0.6
P49755	Transmembrane emp24 domain-containing protein 10	*TMED10*	0.6
P08670	Vimentin	*VIM*	0.6
P16615	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	*ATP2A2*	0.6
Q9UM00	Transmembrane and coiled-coil domain-containing protein 1	*TMCO1*	0.6
O75964	ATP synthase subunit g, mitochondrial	*ATP5L*	0.6
P14859	POU domain, class 2, transcription factor 1	*POU2F1*	0.6
Q9Y3Y2	Chromatin target of PRMT1 protein	*CHTOP*	0.6
Q53H12	Acylglycerol kinase, mitochondrial	*AGK*	0.5
P35580	Myosin-10	*MYH10*	0.5
P36542	ATP synthase subunit gamma, mitochondrial	*ATP5C1*	0.5
P28288	ATP-binding cassette sub-family D member 3	*ABCD3*	0.5
Q9HCU5	Prolactin regulatory element-binding protein	*PREB*	0.5
O43920	NADH dehydrogenase [ubiquinone] iron-sulfur protein 5	*NDUFS5*	0.5
Q9Y3A6	Transmembrane emp24 domain-containing protein 5	*TMED5*	0.5
P08133	Annexin A6	*ANXA6*	0.5

P62805	Histone H4	*HIST1H4A*	0.5
P11233	Ras-related protein Ral-A	*RALA*	0.5
A1L0T0	Acetolactate synthase-like protein	*ILVBL*	0.5
Q9NX63	Coiled-coil-helix-coiled-coil-helix domain-containing protein 3, mitochondrial	*CHCHD3*	0.5
Q9Y276	Mitochondrial chaperone BCS1	*BCS1L*	0.5
P60660	Myosin light polypeptide 6	*MYL6*	0.5
Q96A26	Protein FAM162A	*FAM162A*	0.5
Q8TC12	Retinol dehydrogenase 11	*RDH11*	0.5
Q9Y2H6	Fibronectin type-III domain-containing protein 3A	*FNDC3A*	0.5
P04637	Cellular tumor antigen p53	*TP53*	0.5
Q9Y3E5	Peptidyl-tRNA hydrolase 2, mitochondrial	*PTRH2*	0.5
Q9NPI6	mRNA-decapping enzyme 1A	*DCP1A*	0.5
O43663	Protein regulator of cytokinesis 1	*PRC1*	0.5
O00159	Unconventional myosin-Ic	*MYO1C*	0.5
Q13724	Mannosyl-oligosaccharide glucosidase	*MOGS*	0.5
P43307	Translocon-associated protein subunit alpha	*SSR1*	0.5
Q00765	Receptor expression-enhancing protein 5	*REEP5*	0.5
Q5JRA6	Melanoma inhibitory activity protein 3	*MIA3*	0.5
Q9NTJ5	Phosphatidylinositide phosphatase SAC1	*SACM1L*	0.5
P42858	Huntingtin	*HTT*	0.5
Q07065	Cytoskeleton-associated protein 4	*CKAP4*	0.5
Q13423	NAD(P) transhydrogenase, mitochondrial	*NNT*	0.5
O95573	Long-chain-fatty-acid--CoA ligase 3	*ACSL3*	0.5
O95470	Sphingosine-1-phosphate lyase 1	*SGPL1*	0.5
P18859	ATP synthase-coupling factor 6, mitochondrial	*ATP5J*	0.5
Q7LGA3	Heparan sulfate 2-O-sulfotransferase 1	*HS2ST1*	0.5
Q5HYI8	Rab-like protein 3	*RABL3*	0.5
P45880	Voltage-dependent anion-selective channel protein 2	*VDAC2*	0.5
Q9UH99	SUN domain-containing protein 2	*SUN2*	0.5
Q6PL18	ATPase family AAA domain-containing protein 2	*ATAD2*	0.5
Q9NXF1	Testis-expressed sequence 10 protein	*TEX10*	0.5
P21964	Catechol O-methyltransferase	*COMT*	0.5
Q9C0D9	Ethanolaminephosphotransferase 1	*EPT1*	0.5
Q8N5K1	CDGSH iron-sulfur domain-containing protein 2	*CISD2*	0.5
P46821	Microtubule-associated protein 1B	*MAP1B*	0.5
Q96G21	U3 small nucleolar ribonucleoprotein protein IMP4	*IMP4*	0.5

Q8TEM1	Nuclear pore membrane glycoprotein 210	*NUP210*	0.5
Q9NZ01	Very-long-chain enoyl-CoA reductase	*TECR*	0.5
P78362	SRSF protein kinase 2	*SRPK2*	0.5
Q9H9B4	Sideroflexin-1	*SFXN1*	0.5
Q9P035	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3	*PTPLAD1*	0.5
Q99536	Synaptic vesicle membrane protein VAT-1 homolog	*VAT1*	0.5
P05141	ADP/ATP translocase 2	*SLC25A5*	0.5
P07099	Epoxide hydrolase 1	*EPHX1*	0.5
Q9P0M6	Core histone macro-H2A.2	*H2AFY2*	0.5
P11388	DNA topoisomerase 2-alpha	*TOP2A*	0.5
P07197	Neurofilament medium polypeptide	*NEFM*	0.5
P14735	Insulin-degrading enzyme	*IDE*	0.5
Q00325	Phosphate carrier protein, mitochondrial	*SLC25A3*	0.5
P43304	Glycerol-3-phosphate dehydrogenase, mitochondrial	*GPD2*	0.5
Q9Y6M1	Insulin-like growth factor 2 mRNA-binding protein 2	*IGF2BP2*	0.5
Q9BZF1	Oxysterol-binding protein-related protein 8	*OSBPL8*	0.5
P33121	Long-chain-fatty-acid--CoA ligase 1	*ACSL1*	0.5
P51648	Fatty aldehyde dehydrogenase	*ALDH3A2*	0.5
Q9P2B2	Prostaglandin F2 receptor negative regulator	*PTGFRN*	0.5
Q9UGP8	Translocation protein SEC63 homolog	*SEC63*	0.5
Q9Y282	Endoplasmic reticulum-Golgi intermediate compartment protein 3	*ERGIC3*	0.5
P50336	Protoporphyrinogen oxidase	*PPOX*	0.5
Q9HCE1	Putative helicase MOV-10	*MOV10*	0.5
P20645	Cation-dependent mannose-6-phosphate receptor	*M6PR*	0.5
Q9BU23	Lipase maturation factor 2	*LMF2*	0.5
P58107	Epiplakin	*EPPK1*	0.4
Q8NF37	Lysophosphatidylcholine acyltransferase 1	*LPCAT1*	0.4
P09525	Annexin A4	*ANXA4*	0.4
Q9UHG3	Prenylcysteine oxidase 1	*PCYOX1*	0.4
Q6ZXV5	Transmembrane and TPR repeat-containing protein 3	*TMTC3*	0.4
Q9BXS6	Nucleolar and spindle-associated protein 1	*NUSAP1*	0.4
Q9BUR5	Apolipoprotein O	*APOO*	0.4
Q92522	Histone H1x	*H1FX*	0.4
P07196	Neurofilament light polypeptide	*NEFL*	0.4
Q9NZ45	CDGSH iron-sulfur domain-containing protein 1	*CISD1*	0.4
Q9ULX6	A-kinase anchor protein 8-like	*AKAP8L*	0.4

Q6NUQ4	Transmembrane protein 214	*TMEM214*	0.4
Q709F0	Acyl-CoA dehydrogenase family member 11	*ACAD11*	0.4
Q14254	Flotillin-2	*FLOT2*	0.4
P06396	Gelsolin	*GSN*	0.4
O94901	SUN domain-containing protein 1	*SUN1*	0.4
P56134	ATP synthase subunit f, mitochondrial	*ATP5J2*	0.4
P62745	Rho-related GTP-binding protein RhoB	*RHOB*	0.4
Q16822	Phosphoenolpyruvate carboxykinase [GTP], mitochondrial	*PCK2*	0.4
P53701	Cytochrome c-type heme lyase	*HCCS*	0.4
O75955	Flotillin-1	*FLOT1*	0.4
P04083	Annexin A1	*ANXA1*	0.4
O75477	Erlin-1	*ERLIN1*	0.4
P67812	Signal peptidase complex catalytic subunit SEC11A	*SEC11A*	0.4
P53007	Tricarboxylate transport protein, mitochondrial	*SLC25A1*	0.4
Q9Y394	Dehydrogenase/reductase SDR family member 7	*DHRS7*	0.4
O95563	Mitochondrial pyruvate carrier 2	*MPC2*	0.4
Q96N66	Lysophospholipid acyltransferase 7	*MBOAT7*	0.4
O14662	Syntaxin-16	*STX16*	0.4
Q7Z2K6	Endoplasmic reticulum metallopeptidase 1	*ERMP1*	0.3
O15260	Surfeit locus protein 4	*SURF4*	0.3
P30536	Translocator protein	*TSPO*	0.3
P85298	Rho GTPase-activating protein 8	*ARHGAP8*	0.3
P62834	Ras-related protein Rap-1A	*RAP1A*	0.3