Absolute Quantification of Lysosomal Proteins by Targeted Mass Spectrometry

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematischen-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Peter Robert Mosen

aus

Mainz am Rhein

Bonn, März 2025

Angefertigt mit Genehmigung der Mathematischen-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter/Betreuer: Prof. Dr. Volkmar Gieselmann
Gutachter: Prof. Dr. Christoph Thiele
Tag der Promotion: 08. September 2025
Erscheinungsjahr: 2026

Acknowledgement

I would like to express my gratitude to Prof. Dr. Volkmar Gieselmann, for the opportunity to join his group, entering a new, exciting research field and for serving as my first supervisor. With his critical input, occasionally adopting the role of the devil's advocate, he proved to be a valuable addition in my project. I also want to thank Prof. Dr. Christoph Thiele, as well as Prof. Dr. Joerg Hoehfeld and PD Dr. Marianne Engeser for dedicating time and effort, reading and evaluating my work.

Dominic, I am grateful that our paths crossed. Initially, I joined your new lab for an internship with the intention to learn MS. For this you have been and continue to be the best mentor, when a young (PhD-)student is getting started with LC-MS, learning it from scratch. Beyond your passion for MS, during this time you started your lab combining MS and lysosomal proteomics. With your expertise and enthusiasm for both, you quickly got me excited for your lab's "mission" and you always knew how to form a lab team. The early stages of my/our project, getting started with targeted MS of lysosomes, were challenging. Nevertheless, you have been an incredibly patient and consistently available supervisor, generously sharing ideas and your experience, taking me/us to conferences and encouraging me to explore new approaches instead of adhering to the known, old method and protocol.

Further, I want to say thank you to all the people who were working with me in the lab and who contributed to the individual projects. Your contributions significantly influenced the success of my project and my work would not have been that successful, or as enjoyable without you being around. I am truly grateful to have had you as my colleagues. To Norbert, a special thank you for getting me started with the molecular biology in my project, expressing numerous versions of QconCATs for and with me; and just being around in the institute knowing everything and everyone. Robert, I appreciate your companionship as my bench mate in the MS lab, engaging in hours of discussion until today when it is about data processing/analysis and R programming, as well as anything related to MS instrumentation. Asisa, your scientific journey began in Dominic's lab around the same time. It was much easier to work longer in the evening, knowing that you were also present, diligently working on your project. Also, thank you for being my every-year canoe and carnival-buddy. To all my colleagues, Norbert, Robert, Asisa, Alireza, Shiva, Elham, Fatema, Srigayatri, Sofia, Biswajit, Anne, Pathma, Sara, Yannic, Jasjot and Dhriti, I want to express my appreciation for the enjoyable and fun moments we shared. Thank you Anne, Pathma and Sara for being such fun office desk and back neighbours.

Lastly, I want to express my deep appreciation to my wife Camille, my parents and my family as well as my friends. Thank you for being that supportive and understanding during this long journey.

Table of Content

Abbreviations

Å	Angstrom
AA	acetic acid
ABC	ATP-binding cassette transporter
ACN	acetonitrile
AD	Alzheimer's disease
AGC	automatic gain control
α-MEM	α-minimal essential medium
AMPK	5'-AMP activated protein kinase
AQUA	absolute quantification
AUC	*a*rea *u*nder the *c*urve
BME	β-mercaptoethanol
BORC	BLOC1-related complex
BUP	bottom-up proteomics
BSA	bovine serum albumin
CE/CEO	collision energy/collision energy optimization
CHEVI	class C homologs in endosome-vesicle interaction
CID	collision-induced dissociation
CORVET	class C core vacuole/endosome tethering
CNS	central nervous system
CV	coefficient of variation
DDA	data-dependent acquisition
DIA	data-independent acquisition
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
ECM	extracellular matrix
ER	endoplasmic reticulum
ERT	enzyme replacement therapy
ESCRT	endosomal sorting complex required for transport
ESI	electrospray ionization
FACS	fluorescence-activated cell sorting
FASP	filter-aided sample preparation
FBS	fetal bovine serum
FDR	false discovery rate
FLEXIQuant	*F*ull-*L*ength *E*xpressed Stable *I*sotope-labeled Proteins for *Quant*ification
FT-ICR	fourier-transform ion cyclotron resonance

GAG	glycosaminoglycan
GCl	guanidinium chloride
GO	gene ontology
GSD type II	glycogen storage disease type II
GSK3β	glycogen synthase kinase 3 beta
HD	Huntington's disease
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HOPS	homotypic fusion and protein sorting
HSCT	hematopoietic stem cell therapy
iBAQ	intensity based absolute quantification
ICAT	isotope-coded affinity tag
IGD	in-gel digestion
IP	immunoprecipitation
IPTG	*iso*propyl β-D-1-*th*io*g*alactopyranoside
IS	internal standard
iST	*i*n-*S*top-and-go-extraction *T*ip
iRT	indexed retention time
IT	ion trap
iTRAQ	isobaric Tags for Relative and Absolute Quantitation
LC	liquid chromatography
LC-MRM-MS	liquid chromatography multiple reaction monitoring mass spectrometry
LE	late endosome
LEF	lysosome-enriched fraction
LFQ	label-free quantification
LTL	liver tissue lysate
LSD	lysosomal storage disorder
LYNUS	lysosomal nutrient sensing
MAD	median absolute deviation
MALDI	matrix-assisted laser desorption ionization
MBP	maltose-binding protein
MC	miscleaved
MEF	mouse embryonic fibroblast
MIDAS	MRM-initiated detection and sequencing
MIPA	minimally permutated peptide analog
MP	macrophage
MPS III	Mucopolysaccharidoses type III

MRM	multiple reaction monitoring
MS	mass spectrometry
mTRAQ	mass differential tags for relative and absolute quantification
MWCL	*m*ouse embryonic fibroblast *w*hole *c*ell *l*ysate
m/z	mass-to-charge ratio
mTORC 1	mechanistic target of rapamycin complex 1
Na_2HPO_4	disodium hydrogen phosphate
NaH_2PO_4	sodium dihydrogen phosphate
NCE	normalized collision energy
NMC	non-missed-cleaved
OB	osteoblast
OC	osteoclast
OD	optimal density
OT	Orbitrap
PBS	phosphate-buffered saline
PD	Parkinson's disease
PM	plasma membrane
PNS	post-nuclear supernatant
PrESTs	*Pr*otein *E*pitope *S*ignature *T*ag*s*
PRM	parallel reaction monitoring
PSAQ	protein standard absolute quantification
PSM	peptide-spectrum match
PTM	post-translational modification
x g	relative centrifugal force (rcf)
RpGst	RapiGest
rpm	revolutions per minute
rCV	robust coefficient of variation
rSD	robust standard deviation
RVT	reverse translation
RT	room temperaturef
Q	quadrupole
QconCAT	quantification concatemer
qPCR	quantitative PCR
SDS	sodium dodecyl sulfate
SIL	stable isotope-labeled
SILAC	stable isotope labeling by amino acids in cell culture

SILLPS	stable isotope-labeled lysosomal peptide standards
SLC	solute carrier
S/N	signal-to-noise ratio
SNARE	soluble NSF attachment protein receptors
SPIONs	superparamagnetic iron oxide nanoparticles
SP3	single-pot, solid-phase-enhanced sample preparation
SRT	substrate reduction therapy
Stage-Tip	*st*op-*a*nd-*g*o-*e*xtraction tips
std	standard
STrap	suspension trapping
SWATH	*S*equential *W*indow *A*cquisition of all *Th*eoretical Mass Spectra
TFA	trifluoroacetic acid
TFEB	transcription factor EB
TMT	tandem mass tag
TOF	time-of-flight
Tris-HCl	tris(hydroxymethyl)aminomethane-hydrochloric acid
vATPase	vacuolar ATPase
XIC	extracted ion chromatogram

Summary

Lysosomes are low abundant, membrane-encapsulated cytoplasmic organelles that play a pivotal role in the cellular clearance of various types of extra- and intracellular material including the degradation of biological macromolecules; thereby having major implications for cellular metabolism. Around 340 proteins are known to reside within, or be temporarily associated with the lysosome, making up the lysosomal proteome. Quantitative proteomics experiments utilizing mass spectrometry (MS) play a crucial role in gaining comprehensive insights into the organelle's biology and in understanding its involvement and alterations in various diseases.

However, most studies on lysosomes, or such covering lysosomal proteins using MS, have primarily resulted in relative quantitative data, with researchers investigating changes in protein levels between different samples and experimental conditions. Absolute quantities of lysosomal protein remain largely unknown, although such data can be considered as the most desirable/universally applicable outcome of a quantitative experiment. This study utilized the quantification concatemer (QconCAT) strategy to create an absolutely quantified internal standard for 144 lysosomal proteins, covering the majority of the mouse lysosomal core proteome in an MS experiment. Unlike relative quantitative data, the acquired absolute quantitative data allow not only for the calculation of absolute quantities, but also for the comparison of lysosomal protein amounts between sample types and conditions. In addition to the numerous analytical advantages offered by the internal standard applied, absolute quantitative data hold the promise to enhance our understanding of the molecular relationship between different proteins in and across various samples. To further overcome current shortcomings and limitations associated with untargeted MS experiments and lysosome enrichment, limiting the compatibility of certain sample types with MS analyses, targeted MS approaches for the direct study of lysosomal proteins from whole cell and tissue lysates were developed. As a result, a systematic and unbiased quantitative study across various cell/tissue types of varying complexities, independent of any lysosome enrichment/isolation procedures is now possible.

For the investigation of samples with varying input amounts and sample complexity, first a robust and universally applicable sample preparation workflow was established using the MS-compatible surfactant RapiGest. Here, the peptide/protein concentration emerged as a critical parameter in the sample preparation workflow, influencing the precipitation behavior of the surfactant. Further, for the sample preparation of low-concentrated samples a workaround involving a trigger protein for co-precipitation of the surfactant was presented. In an initial, label-free comparative study quantifying lysosomal proteins from various complex sample lysates using data-independent acquisition and parallel reaction monitoring, it became evident that a targeted MS acquisition strategy such as parallel reaction monitoring, surpasses data-independent acquisition-MS in detecting and accurately quantifying subtle changes of low-abundant (lysosomal) proteins within highly complex sample backgrounds such as tissue lysates. For the absolute quantification of

144 lysosomal proteins from any cell/tissue type of mouse origin a targeted MS assay employing multiple reaction monitoring was developed and the stable isotope-labeled, absolutely quantified internal standard was applied. The third study provides absolute quantitative data on 143 lysosomal proteins in one cell line, lysosome-enriched fractions and four primary cell types, enabling the determination of individual lysosomal protein copy numbers. In mouse embryonic fibroblasts, the lysosomal core proteome composition exhibited a dynamic range spanning three orders of magnitude. The spatial distribution of individual proteins and protein groups between lysosomal and non-lysosomal compartment revealed feature-specific localization patterns, with substantial proportions of membrane proteins and partially hydrolytically active proteins localized in non-lysosomal compartments. Across the different cell types investigated, variations in the abundance of distinct lysosomal protein classes and in proteome composition underscored the functional specialization of individual cell types, with primary cells showing elevated levels of hydrolases/luminal proteins. Moreover, unique expression dynamics of functional protein classes as well as selected proteins within these classes highlighted cell type-specific lysosomal adaptations tailored to their specific functions. qPCR analysis of 51 lysosomal proteins demonstrated that this cell type-specific regulation of the proteome is - with only a few exceptions - predominantly driven by transcriptional control.

1. Introduction

1.1. The lysosome – central degradative organelle and signaling hub of the cell

The lysosome, with its name derived from the Greek word lysis (dissolution, destruction) and soma (body) is a degradative component of the cell. It catabolizes biological macromolecules like proteins, lipids, carbohydrates, and nucleic acids [1, 2], ranging from single molecules to larger structures such as protein complexes [3], organelles [4] and whole cells [5, 6]. Thereby the lysosome plays a pivotal role in the cell's endeavour to maintain cellular homeostasis, which characterizes its biologically stable and functional status [7]. Present in almost all eukaryotic cells, lysosomes are membrane-enclosed, cytoplasmic organelles, which receive lysosomal substrates - depending on the spatial origin of the macromolecule - by multiple routes of transport. While extracellular material is being transported by endocytic vesicular mechanisms, intracellular material reaches the lysosome via autophagosome vesicular transport (Figure 1). A special type of lysosomal substrate delivery is non-vesicular-based chaperone-mediated autophagy, which allows a direct transport of unfolded proteins across the lysosomal membrane into the lysosomal lumen [2]. In the acidic lysosomal lumen (pH 4.5-5.5) which is established and sustained through a membrane-bound vacuolar ATPase-proton pump complex [8], 60-80 compound class-specific acidic hydrolases (proteases, lipases, glycosidases and nucleases) break down the macromolecules to their low(er) complex molecular units. Generated end products are transported back to the cytosol via transporters and are further re-utilized in cellular processes [9].

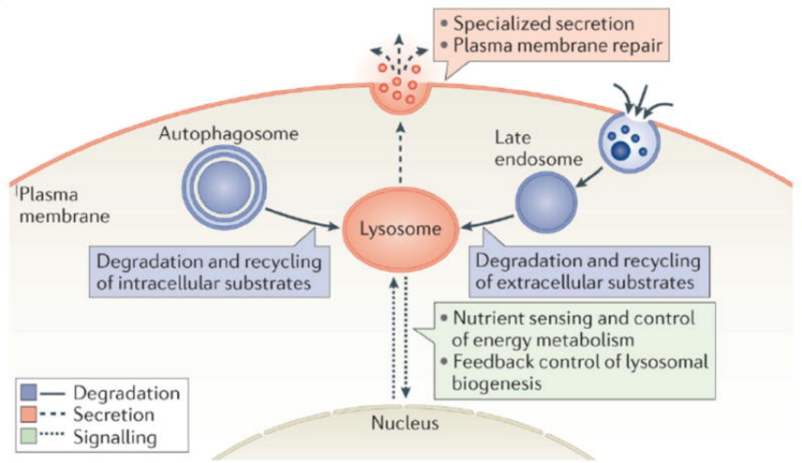


Figure 1: Key functions of the lysosome and its integration into the cellular framework. Intra- and extracellular material is transported via autophagosome vesicular and endocytic vesicular mechanisms to the lysosome. Vesicular cargo fuses, as autophagosomes or late endosomes, with the lysosome, where macromolecules are degraded into low(er) complex molecular units, and recovered catabolic products can be transported back to the cytosol (not shown) [2]. Lysosomes can alter their position, moving closer to, and fusing with the plasma membrane. This enables them to either release their organellar content (luminal hydrolases and functional substrates) into the extracellular space or to engage in plasma membrane repair and -remodeling [10]. Lysosomes function as "signaling hub" of the cell. Sensing nutrient availability and other environmental stimuli, such as the energy status of the cell, lysosomes take part in regulating the delicate equilibrium between growth-promoting anabolic and catabolic activities within the cell (not shown). Further, a direct transcriptional control of lysosomal activity and biogenesis, as well as of the autophagosomal-lysosomal pathway is possible [11]. Figure from [2].

The view on and our understanding of the lysosome has changed and extended since its discovery [12]. Initially considered as a static, constitutive active organelle which operates detached from other cellular processes, research has shown that the lysosome is a dynamic and highly regulated organelle with implications that go way beyond its catabolic function and role in the cell. Lysosomes are of critical importance in major biological processes such as metabolic sensing and signaling [13], gene regulation [14, 15], antigen presentation [16], plasma membrane repair [17], exocytosis [10, 18], as well as cell adhesion and migration [19]. For this purpose the lysosome can engage in interactions with multiple cellular compartments (plasma membrane and extracellular space [18]; endoplasmic reticulum (ER), peroxisome and mitochondria [20]; Golgi apparatus [21]) and vesicles, such as autophagosomes [22] which are either of fusogenic or non-fusogenic nature (Figure 1). For example, the interaction between mitochondria and lysosomes enables a direct, bidirectional regulation of organelle dynamics through contact sites which can influence the acidification and function of the lysosome [23, 24]. Furthermore, lysosomes participate in the alteration of cellular function and biological processes, adjusting and influencing major regulators of cell metabolism and growth such as the mechanistic target of rapamycin complex 1 (mTORC 1) [25], glycogen synthase kinase 3 beta (GSK3β) [26] or 5'-AMP activated protein kinase (AMPK) [27]. By transient re-localization, attachment and activation at the lysosomal surface, kinases and kinase-complexes integrate a variety of signals on the cell's condition, e.g. nutrient and energy availability, cell growth and metabolism, stress etc., via their signaling pathways and initiate further downstream signaling events. mTORC1 is part of the lysosomal nutrient sensing (LYNUS) machinery, a multiprotein-complex assembly formed in nutrient-rich conditions on the lysosomal surface [2], which senses lysosomal and cytosolic amino acid levels [25, 28, 29]. As a result, anabolic processes of the cell such as protein biosynthesis and cell proliferation are promoted while catabolic processes (e.g. autophagy) are limited [11] (Figure 1). But also a more direct control of lysosomal activity via the transcription factor EB (TFEB), a member of the microphthalmia/transcription factor E family [30], was discovered in the case of nutrient scarcity. Due to the absence of TFEB phosphorylation, TFEB translocates to the nucleus [31], where it promotes transcription of 471 direct TFEB targets increasing lysosomal function and biogenesis. Proteins regulated do not only comprise lysosomal proteins but also those associated with lysosomal biogenesis as well as autophagy [32] (Figure 1). In addition to nutritional deprivation, lysosomal, ER, mitochondrial, and infectious stress can also trigger post-translational modifications on TFEB. These modifications are not exclusively limited to phosphorylation. For instance, acetylation, ubiquitination, and other modifications on TFEB have been identified, influencing localization, transcriptional activity, or stability of TFEB [33]. With these mechanisms, lysosomal activity is closely coupled to the nutrient availability and the overall physiological condition of the cell.

1.2 Lysosomal storage disorders

Lysosomal storage disorders (LSDs) summarize a group of over 70 inborn, metabolic disorders, each presenting its individual pathophysiology and disease progression. In all of these disorders, lysosomal malfunction leads to a gradual accumulation of unprocessed lysosomal substrates and/or metabolic intermediates in the lysosome. Generally considered as a rare disease when considering the occurrence of individual LSDs (e.g. 1 in 40,000 live births for Pompe disease), the overall cumulative incidence is estimated to be around 1 in 5,000 live births [34]. The impairment in lysosomal degradation and recycling first results in the dysfunction of the lysosome, but entails a further impairment of lysosome-associated (e. g. autophagy, vesicle trafficking) and other critical cellular processes subsequently leading to the death of the cell. Most LSDs are caused by monogenic mutations in genes encoding predominantly for the diverse group of lysosomal luminal hydrolases (~ 50 enzymes). However, membrane proteins and transporters can also be affected [35], resulting in both cases in the accumulation of storage material. Due to the diverse biochemical types and structures of storage material as well as the cell type-specific abundance, the effect of storage material on the lysosomal function differs widely which might explain the individual pathophysiology and disease progression of LSDs. Despite the heterogeneous disease character, a common feature most LSDs share is neurodegeneration, causing severe central nervous system (CNS) symptoms upon disease progression. Accompanied are these neurological impairments often by an enlargement of abdominal organs (visceromegaly). As a result of the perpetual disease progress multiple organs get affected (multisystemic phenotype), whereas one organ is predominantly impaired driving disease progression and showing organ-specific disease characteristics [35]. In addition to the direct connection of lysosomal protein malfunction and LSDs, lysosomal proteins have been associated with other, more prevalent conditions. This includes neurodegenerative diseases such as Alzheimer's, Parkinson and Huntington's disease [36-38], but also in cancer lysosomal alterations have been identified [39]. With a mostly early disease onset, an untreated LSD leads to the premature death in early childhood.

Treatments are available for only some of the LSDs, however the treatment options can typically only mitigate and manage symptoms rather than cure the disorder. Available treatments follow three approaches and include: enzyme replacement therapy (ERT; intravenous injection of a recombinant, functional lysosomal enzyme), substrate reduction therapy (SRT, limiting the accumulation of substrate through inhibition of metabolite-converting enzymes) and chaperone therapy (enhancing enzyme stability and maximizing lysosomal catalytic performance of the endogenous mutant enzyme variant) [35]. Other treatments like hematopoietic stem cell therapy (HSCT) and gene therapy-based approaches also pursue the idea of enzyme substitution and further enzyme correction [34]. In hematopoietic stem cell therapy the deficient enzyme variant is replaced and supplemented by the

functional enzyme variant produced by the transplanted stem cell(s). Gene therapeutic approaches involve the correction of the genetic defect itself using viral vectors and other vehicles for DNA transfer. Gene correction can be achieved either *ex vivo* (in vitro genetic correction in extracted patient cells followed by re-transfer) or *in vivo* (intravenous administration of a viral vector loaded with a therapeutic, functional gene)[40]. As examples of LSDs, the following subsections will provide a brief overview of two storage disorders: mucopolysaccharidosis type III (MPS III) and glycogen storage disease type II (GSD II, Pompe disease). These two disorders will be introduced and discussed in the context of their molecular basis, pathophysiology and relevance to lysosomal function.

1.2.1 Pompe disease

GSD type II, or acid maltase deficiency is an autosomal recessive disease which is caused by the deficiency of acid α-glucosidase (*GAA*), a lysosomal hydrolase responsible for the breakdown of glycogen [41]. Glycogen, a complex and highly branched carbohydrate composed of glucose molecules, serves as a storage form of energy and is predominantly stored in the liver and muscles of mammals [42]. In GSD II, glycogen storage material accumulates in the lysosomes of multiple tissues/organs, where primarily skeletal and cardiac muscles are affected and damaged [43]. The pathology extends beyond the muscle cells and involves accumulation of glycogen in motoneurons [44, 45], resulting partially in a neuromuscular pathology in the progressing disease (respiratory and cardiac dysfunction) [43]. Depending on the age of disease onset, which can occur at any time [35], a wide phenotypic spectrum of symptoms is known. A shared characteristic is the progressive muscle hypotonia accompanied by a gradual loss of motor functionality and finally respiratory failure [43]. While at a late disease onset (juvenile/adult) a slow progression of skeletal myopathy is described (cardiac muscle is spared), the more common, and severe, infantile disease variant is accompanied by hypertrophic cardiomyopathy (pathological thickening of the left heart ventricle) leading to premature death within the first two years of life [46].

1.2.2 Mucopolysaccharidosis type III (MPS III)

MPS type III, or Sanfilippo syndrome is a neurodegenerative LSD from the group of mucopolysaccharidoses, comprising four subtypes MPS III A-D, all characterized by an autosomal recessive inheritance. Common disease pathology of MPS III subtypes is the lysosomal accumulation and urinary excretion of heparan sulfate [47], a sulfated glycosaminoglycan (GAG) variant, caused by mutations in genes *SGSH* (MPS type III A), *NAGLU* (MPS type III B), *HGSNAT* (MPS type III C) or *GNS* (MPS III D) [35, 48]. Heparan sulfate consists of multiple disaccharide units, with each unit being made up of glucuronic acid linked to (N-acetyl)glucosamine. These units can undergo modifications, resulting in various types of heparan sulfate differing in composition and

distribution of modified disaccharide units. Heparan sulfates are further an integral component of larger proteoglycan structures playing essential roles on the cell surface and in the extracellular matrix (ECM) in most animal tissues. Proteoglycans, which are glycoproteins, are involved in a wide array of cellular functions and processes, such as cell migration, motility and adhesion, vesicle secretion, endocytosis but also provide structural support to basement membranes, glial and neuronal cells [48]. Whereas under normal cellular conditions heparan sulfates undergo successive degradation in the lysosome which proceeds in a strictly sequential manner [49, 50], under pathological conditions this catabolic process is impaired. Depending on the MPS III subtype the different enzymes and thereby different steps in the heparan sulfate catabolism are affected, resulting in the production of sulfated glycosaminoglycan intermediates of different chain length [49, 50] (Figure 2).

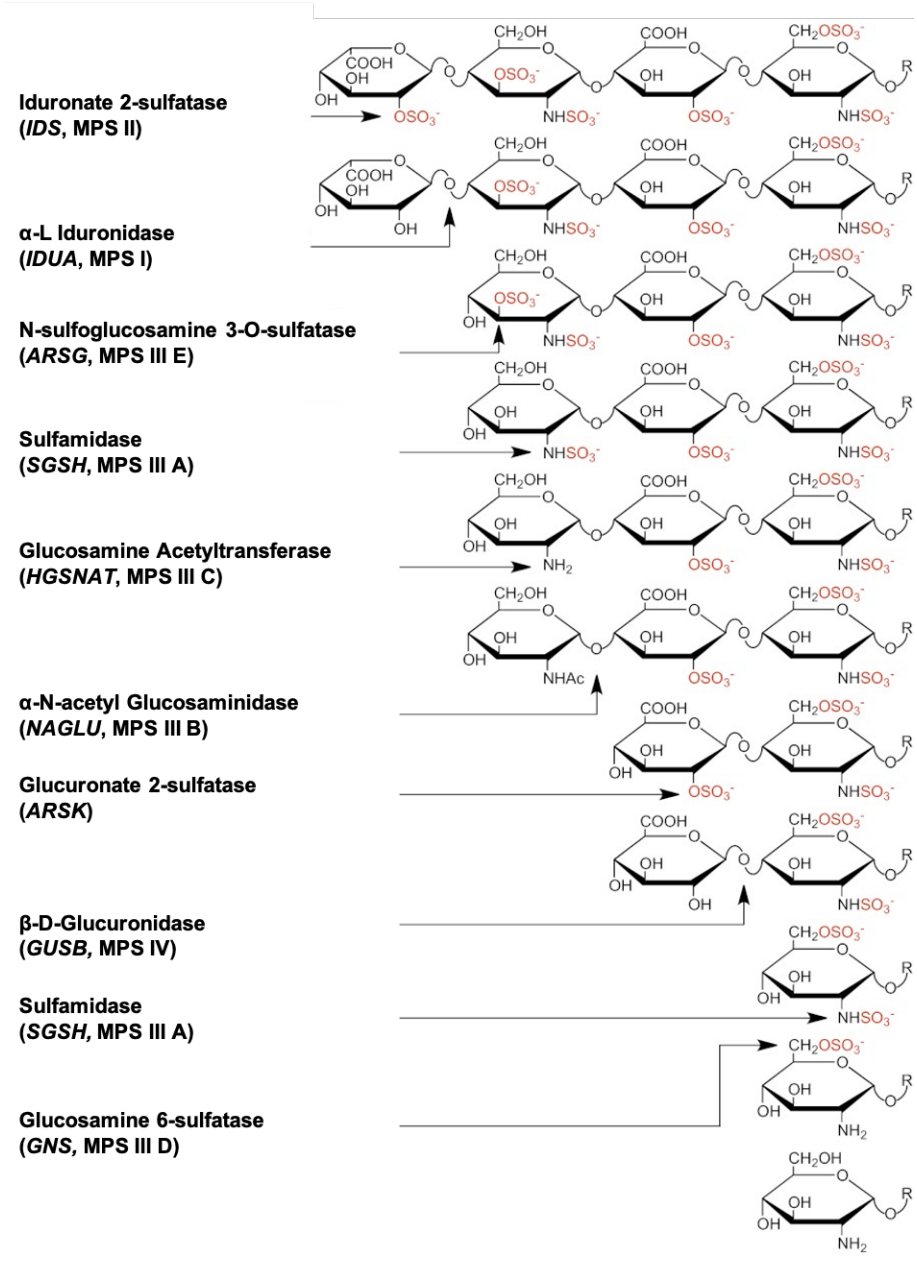


Figure 2: Schematic representation of the heparan sulfate catabolic pathway. Heparan sulfate reaches the lysosome through the endosomal pathway where it undergoes stepwise degradation. Proceeding in a strictly sequential manner, a dysfunction in any of the nine involved enzymes leads to the accumulation of sulfated glycosaminoglycan intermediates causing mucopolysaccharidoses (MPSs). Names of affected genes and mucopolysaccharidoses subtypes are indicated. Figure from [49].

Symptoms for the different subtypes of MPS III are similar [35]. Characteristic is an early onset (late infantile to juvenile age [35]), and progressive, severe degeneration of the CNS which is accompanied by mild somatic symptoms [48]. Although the neurodegeneration does not manifest with immediate symptoms, it becomes evident during the second phase of the disease (neuroregression phase), beginning around the age of four years. This phase is marked by a progressive intellectual disability, abnormal and aggressive behavior, feeding/eating difficulties and motor deterioration, cognitive impairment, dementia, hyperactivity, speech loss as well as organ malformation (e.g. heart) and musculoskeletal abnormalities. In the terminal phase, deterioration of fundamental motor functions such as walking and autonomic functions such as swallowing continues, finally leading to death due to respiratory complications and cardiorespiratory failure [51] in the second or third decade of life [48, 50].

1.3. Study of the lysosomal proteins and the lysosomal proteome

Traditionally, most research on the lysosome and lysosomal proteins has been conducted on a small(er) scale, employing either biochemical techniques such as Western blot, enzyme assays and immunohistochemistry or microscopy using pH sensitive probes and fluorophore-coupled inhibitors [9]. However, the idea of investigating one protein, or a smaller subset of lysosomal proteins that are representative of the whole organelle, falls short in depicting a full organelle's complexity and dynamics [52]. The tight involvement of the lysosome in diverse cellular processes and the multiple points of interaction with other cellular structures as briefly outlined in Chapter 1.1, calls for a comprehensive analytical approach covering as many lysosomal proteins as possible. Modern mass spectrometry (MS) approaches enable an unbiased, comprehensive view on the proteome, which can be understood as the "set of proteins expressed by an organism" [53]. Muntel et al. reported the quantification of more than 10,000 proteins in a six hour single shot MS experiment from unfractionated tissue samples [54]. Messner et al. reduced the experimental time of one single shot experiment to five minutes and was still able to quantify more than 4,000 protein groups from a human cell line using ultrafast scanning MS [55]. Even though these are extreme examples of maximized MS performance, it shows that modern MS approaches enable large-scale proteomic studies with respect to sample depth as well as sample numbers/throughput.

Over time, various MS techniques and approaches have also been applied to investigate the lysosome and its proteins. Whether employed for the quantification and identification of lysosomal proteins [56], or the characterization of lysosomal proteins/protein classes [57], MS has played an important role in extending and shaping our view on the lysosome. From public databases, more than 900 proteins can be linked to the lysosome based on associated Gene Ontology (GO) terms such as 'lysosomal' or 'lysosome'. For a subset of 341 proteins, a „specific lysosome-related biological function" [9] and/or an involvement was confirmed based on an experimental, orthogonal method providing a manually curated high confidence list of lysosomal proteins [9]. This list of

proteins comprises beside the ~70 luminal, acid hydrolases more than 100 integral (trans-) membrane proteins which ensure the structural integrity of the organelle, enable the transport of molecules across the membrane and serve as an anchor point for organelle-organelle or organelle-cellular compartment interaction (Figure 3). Because of their persistent localization and function, whether in the lumen or membrane, these proteins constitute a kind of 'lysosomal core proteome' [58]. Closely related to the fraction of (trans-) membrane proteins and their functional involvement are another ~120 lysosome-associated proteins/complexes [9]. Residing in the cytosol these proteins localize to the lysosomal surface/membrane, where they can temporarily interact with other proteins and can form multimeric protein structures further participating in various cellular processes [2]. Some examples of proteins/protein complexes are: vATPase complex, an ATP-dependent proton pump complex establishing the acidic environment in the lysosome which consists of a cytosolic-(v1) and lysosomal membrane-bound (v0) multimeric protein region associating [8, 59]; formation of the RagGTPases-Ragulator complex which interacts with the mTOR complex 1 (mTORC1) [60] and is required for nutrient sensing, metabolic signaling as well as adaptation [13, 61]; BORC (BLOC1-related complex) which is involved in lysosomal movement and positioning [62]; SNARE (soluble NSF attachment protein receptors), CORVET (class C core vacuole/endosome tethering) and HOPS (homotypic fusion and protein sorting) which are essential in lysosomal membrane fusion events e.g. with other organelles [63]; as well as the ESCRT (endosomal sorting complex required for transport) machinery which maintains membrane integrity of the endolysosomal system through membrane repair activity [64] (Figure 3).

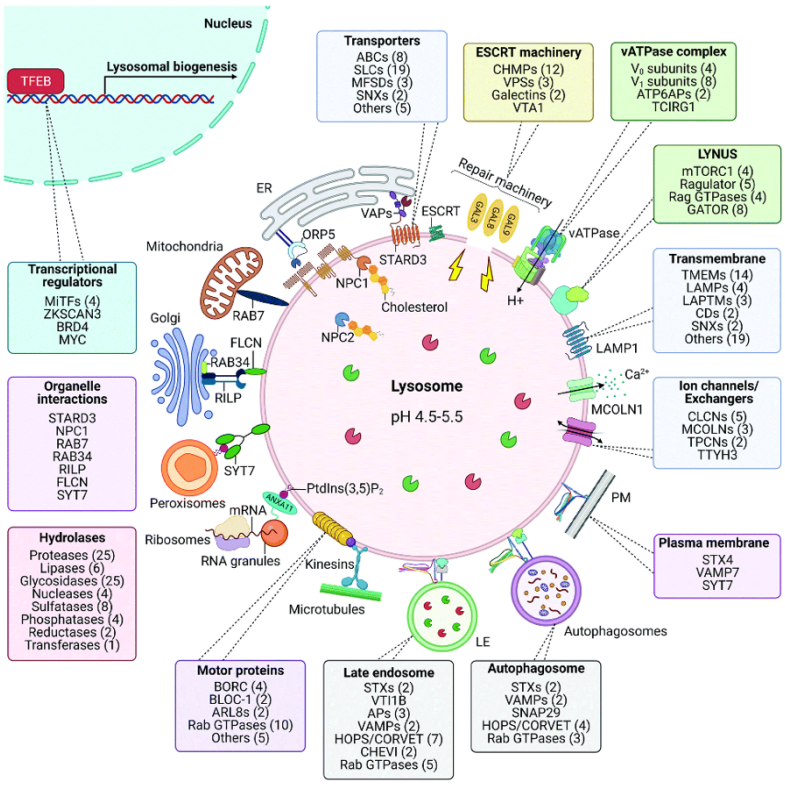


Figure 3: Draft of the lysosomal proteome. The lysosomal proteome can be categorized into three protein groups: luminal, membrane and cytosolic, membrane-associated (temporarily interacting) proteins. Individual proteins are assigned to classes, based on their function and/or protein complex affiliation. Localization, as well as the number of members per class is shown. Abbrev.: CHEVI: class C homologs in endosome–vesicle interaction; CORVET: class C core vacuole/endosome tethering; ESCRT: endosomal sorting complex required for transport; HOPS: homotypic fusion and protein sorting; LE: late endosomes; LYNUS: lysosomal nutrient sensing machinery; PM: plasma membrane. Figure from [4].

1.4. Mass spectrometry as a tool to study the proteome

With the progress made over the last two decades on developing modern mass spectrometers [65, 66] and associated bioinformatics algorithms [67, 68], MS has become a valuable tool for molecular and cellular biologists. Mass spectrometers operate based on a basic principle: accurately measuring the mass-to-charge ratio (m/z) of charged molecules, e.g., peptide ions. However, before accurate mass measurements can take place, analytes must be ionized and transferred to the gas-phase [69, 70]. While multiple strategies have been developed to achieve gas-phase ionization, electrospray ionization (ESI) [71] and matrix-assisted laser desorption ionization (MALDI) [72] have established themselves in the field of MS-based proteomics. This stems from the 'soft' nature of the ionization process in MALDI and ESI, meaning ionization of large and fragile biomolecules such as peptides or proteins takes place without fragmentation [73] ensuring that the molecular integrity is preserved during ionization. After ionization, the analytes enter the mass spectrometer, where mass analysis takes place and m/z values of the analyte ions are determined. At the same time, the number of ions detected is recorded as an intensity signal [70, 74]. For separation and determination of the m/z different mass analyzers exist, which differ in the principle of separation and detection of ions, such as the Fourier-transform ion cyclotron resonance (FT-ICR), the time-of-flight (TOF), the ion trap (IT), the quadrupole (Q) [73] and the Orbitrap (OT) [65]. Tightly connected to their operational principles are also the instruments' individual advantages and limitations, e.g., acquisition speed, limit of detection, mass range or resolution.

To overcome mass analyzer-specific limitations and to leverage the mass analyzer-specific advantages, often two or more mass analyzers are combined in a single instrument, creating the so-called hybrid mass spectrometers. In hybrid mass spectrometers, tandem MS experiments involve two (or more) consecutive stages of mass analysis which are performed in the different mass analyzers [73]. In an initial step, at the precursor level, a mass spectrum of the intact ion analyte (precursor ion) is recorded. Following the isolation and fragmentation of the intact ion analyte, a mass spectrum of generated fragment ions arising from the precursor fragmentation is acquired. This spectrum provides valuable information about the structural composition of the intact peptide [70, 74]. Hybrid instruments enable comprehensive, multi-step proteomic analyses, elucidating structural details of the molecule or performing multiplexed quantification [73], which makes MS an ideal tool for studying proteomic heterogeneity. Understanding such details is key to start comprehending the complexity of the proteome, which is known to be notably larger than the complexity of the genome when considering proteoforms arising from alternative splicing, post-translational modifications (PTMs), etc. [73, 75]. Proteins can be studied by MS, either using a top-down or a bottom-up proteomics (BUP) strategy. Top-down proteomics analyzes individual, intact proteins, which is crucial for direct analysis of proteoforms and other molecular complexity features of a protein, since these are preserved and visible in the mass spectrum [76]. However, top-down proteomics suffers from

multiple limitations, such as the need for a delicate and laborious sample preparation, achieving low(er) throughputs than its counterpart BUP [77]. In contrast to that, BUP encompasses the analysis of peptides generated from protein analytes by proteolytic digestion. The peptides serve as the foundation for identifying and quantifying proteins and in combination with other bioanalytical techniques such as liquid chromatography, the analysis of complex protein mixtures is possible. This peptide-centric strategy for the study of the proteome will be discussed in further detail below.

1.4.1 Bottom-up MS proteomics using liquid chromatography

BUP is the most common MS-based approach used to study the proteome. Combining two powerful analytical techniques, liquid chromatography (LC) and MS, BUP enables the detailed and quantitative study of thousands of proteins simultaneously from complex samples [78]. In a modern, generic BUP experiment, proteins are first extracted from cells or tissue material using a combination of physical and chemical lysis. Proteins present in a crude mixture are then digested into peptides by a protease such as trypsin, which cleaves after amino acids lysine (K) and arginine (R), and generate a highly complex peptide mixture (Figure 4). Despite the increase in complexity, the cleavage specificity of the protease ensures that resulting peptide analytes will be similar in certain physicochemical properties, e.g. basic termini, which facilitates LC-MS analysis and subsequent data analysis [79]. Optionally to the direct processing of proteins to peptides, as described above, enrichment and fractionation as well as depletion strategies can be implemented into the sample preparation workflow. Applied either at protein or peptide level these additional steps in sample preparation aim to reduce the complexity arising from the proteolytic digestion, ultimately contributing to a deeper coverage of the proteome [80-86]. The obtained proteolytic digest of the entire proteome or the fraction(s) yielded in case of a fractionation, enrichment or depletion experiment, are then injected onto a liquid chromatography system which is coupled to an MS instrument. In a standard BUP proteomics experiment as described here, peptide separation typically entails employing a reversed-phase liquid chromatography setup, although alternative liquid/stationary phase combinations are feasible in a LC-MS experiment. In reverse phase liquid chromatography, a proton-rich mobile phase solvent ensures protonation of peptides which are separated over time by reverse-phase LC based on their hydrophobicity, reducing the complexity for the upcoming MS analysis [74]. Protonated peptides are ionized and transferred to the gas-phase by electrospray ionization (ESI), which generates multiply-charged peptide ions from highly charged liquid droplets by simultaneously applying high-voltage, heat [87] and vacuum [71]. Peptide ions that have entered the mass spectrometer are directed by electric fields in near full-vacuum conditions and MS data is acquired.

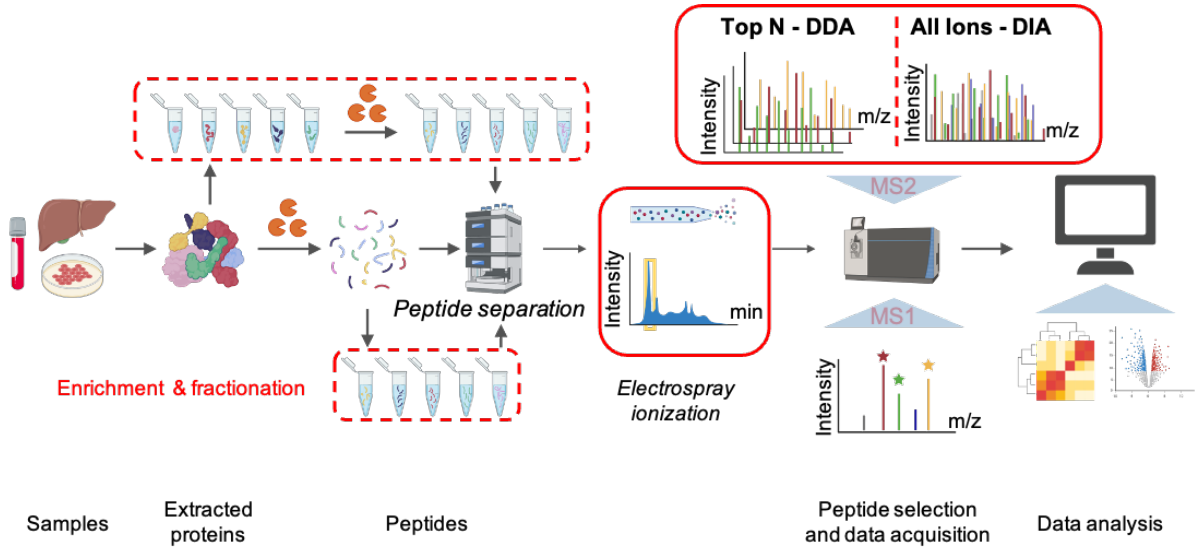


Figure 4: Workflow of a typical bottom-up MS proteomics experiment. Proteins extracted from diverse sample types can directly be proteolytically digested into peptides. Alternatively, various methods for protein or peptide enrichment and fractionation can be implemented. Generated peptides are separated through liquid chromatography, ionized and subjected to MS analysis. Data are acquired at the MS1 and/or MS2 level (intact precursor and fragment ions, respectively) across the chromatographic gradient. Depending on the data acquisition approach selected, the MS2 spectra acquisition differs, resulting in either a) a sequential acquisition of individual fragment ion spectra for the most abundant precursors (DDA) or b) highly multiplexed fragment ion spectra combining fragment ion spectra of all precursors (DIA). Acquired data is pre-processed and can be analysed using dedicated software tools and algorithms. Abbrev.: DDA: data-dependent acquisition; DIA: data-independent acquisition. Figure was created with BioRender.

For data acquisition two modes, data-dependent acquisition (DDA) and data-independent acquisition (DIA) can be distinguished (Figure 5). In data-dependent acquisition (DDA), first the m/z of all the peptide precursors present at a given timepoint of the LC gradient are determined with an MS1 scan. Based on this scan, the top N most abundant precursors are selected and sequentially directed to the collision cell for fragmentation. Following the peptide precursor fragmentation, an MS2 spectrum containing the m/z information for each of the generated fragment ions is recorded [74]. However, this traditional way of precursor selection and data acquisition suffers from one main drawback: while covering most of the abundant proteins (usually the top 10-20 precursors are selected [88]), stochastic sampling of lower abundant proteins leads to reduced reproducibility for proteins in the medium-to-low abundance range, resulting in higher missing value rates [89]. Therefore, in recent years data-independent acquisition (DIA) strategies have gained popularity. Here, all precursor ions present are fragmented, regardless of their abundance. For this purpose, the MS1-scan range is divided into large and slightly overlapping m/z windows, normally encompassing around 25 Thomson or m/z units [90]. Within these windows, all the precursor ions are isolated in parallel and fragmented, resulting in complex, highly multiplexed MS2-scans (hybrid MS2-spectra), containing the fragments of all co-isolated peptide precursors in the given m/z range. Due to the continuous cycling of the fixed

precursor isolation windows, the full m/z range is covered, generating extensive datasets with a theoretically complete MS2 data coverage [91].

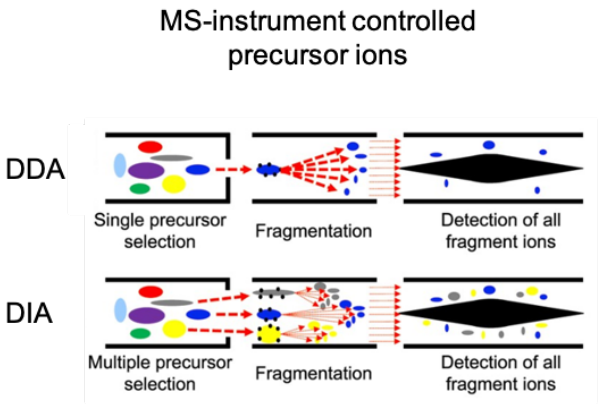


Figure 5: Common operational modes of an MS instrument. In DDA, single precursor ions are selected by the instrument (software) for isolation, precursor fragmentation, and fragment ion data acquisition based on their abundances. In contrast, in DIA, the instrument isolates and fragments all precursor ions within a specified mass range (m/z window) followed by one fragment ion scan. Abbrev.: DDA: data-dependent acquisition; DIA: data-independent acquisition. Figure from [92].

Multiple benchmarking experiments have demonstrated that DIA strategies surpass DDA, by providing better quantitative reproducibility, increased sensitivity as well as superior specificity and accuracy [54, 93]. These benefits, as well as novel software developments that will be discussed briefly in the next section, explain the recent preferential adoption of DIA-based strategies in the proteomics field.

1.4.2. Peptide fragmentation and identification

An unambiguous peptide identification is essential to infer the protein composition of the original sample. Since an unambiguous peptide identification is (almost) never possible from the MS1 scan i.e the precursor m/z alone, precise determination of the amino acid sequence is required. For example, two peptides can have the exact same *m/z* but a distinct sequence, which is known as peptide isomers, and these would remain indistinguishable at precursor ion level. Therefore, to infer the peptide sequences, fragmentation is needed. During fragmentation, the intact peptide precursor ion is fragmented into a series of smaller N- or C-terminal fragments [74]. Over the years, various fragmentation strategies have been developed with the aim of achieving sufficient and efficient fragmentation for different applications, e.g. the analysis of diverse PTMs [94, 95]. Different fragmentation strategies tend to break the precursor ions at specific chemical bonds over the peptide backbone. Based on the nature of the chemical bond dissociated, a series of resulting fragment ions is released. These ion series are called a-/x-fragments if the bond between Cα-CO is cleaved, b-/y-fragments if the peptide bond or C-N bond is cleaved, or c-/z-fragments if the N-Cα-bond is cleaved [96]. In this work, collision-induced dissociation (CID) strategies have been used. CID strategies lead to peptide dissociation preferentially at the peptide bond (C-N bond) when the intact precursor ion collides with a neutral gas, e. g. nitrogen or helium [97]. To unambiguously identify peptides and determine their sequences, experimentally detected fragment ion m/z values belonging to either the

b- or y-ion series, are compared against theoretical b- or y-fragment ion series, respectively their calculated m/z values. These theoretical fragment ion series are calculated for all potential peptides within the relevant precursor m/z range [98].

To automatize and speed up the process of peptide identification, computational approaches have been developed and are nowadays widely used, with manual spectral inspection only required for verification of crucial candidates. Two main strategies need to be distinguished in terms of software-assisted spectral interpretation: *de novo* peptide sequencing [99], which tries to derive the sequence of a peptide solely based on the observed m/z difference within a fragment ion series, or database searching strategies, which use a probabilistic-based matching approach to pre-defined sequences in a database [98, 100]. For the identification of peptides by database-search strategies, the experimental information acquired on precursor m/z and its associated fragment ion m/z of unknown identity is being matched to an *in silico* generated precursor and fragment ion m/z for peptides of known identity. For each peptide match or identification, a score is assigned, which reflects the chance of the identification being derived from a random peptide match, therefore controlling for false positive peptide identifications [98, 101]. Similar to the principle of the m/z value-based peptide matching is the spectral library search approach. In the latter, instead of matching theoretical m/z values generated from a protein sequence database, the matching process involves direct comparison of an experimental MS2 spectrum with a reference MS2 spectrum [98, 101]. Although spectral library search approaches were originally developed for DDA data [102], nowadays their use is mostly limited to the growing field of DIA-based studies. As discussed earlier, DIA-MS acquires complex MS2 spectra containing fragment ion series of multiple co-fragmented peptides [91]. As a result, the chimeric nature of DIA-derived MS2 spectra makes it computationally challenging to process and analyze such data. Therefore, the early DIA-based studies were searched against reference spectral libraries generated by pooling thousands of clean and pre-identified DDA-derived spectra. However, with the advent of neural networks, spectra can nowadays be predicted *in silico* for virtually any (tryptic) peptide, even if a DDA-spectrum has never been acquired for it, enabling library-free, fast and reliable DIA data analysis [103, 104].

1.5. MS-based proteomics of lysosomes

The scarcity of lysosomes within cells presents a significant hurdle in MS-based lysosomal proteomics. With only around one hundred up to multiple hundred lysosomes per cell as reported for COS-7 cells [105] and an approximately 9x lower occupancy volume of the lysosome in the cell compared to mitochondria (10-20% occupancy volume per cell depending on the cell type) [106], lysosomes make up only a sub-fraction of the cellular content. In addition to the low abundance of lysosomes itself, lysosomal proteins contribute considerably less than 1% to the total cell protein mass (0.2% in HeLa and 0.1% in neurons) [107, 108]. Beyond the challenge posed by low

abundance, the wide dynamic range of lysosomal protein expression which is estimated to span up to four orders of magnitude [56], further complicates the reproducible identification and quantification of lysosomal proteins from whole cell lysates (WCLs). Particularly the untargeted, abundance-dependent DDA acquisition strategy suffers from the sample complexity [89]. While the detection of medium and low abundant proteins is often not reproducible mainly the high abundant, luminal lysosomal proteins are detected. Low abundant lysosomal proteins for example membrane proteins are detected less frequently or even remain uncovered. To overcome the limitations in sensitivity, lysosomes can be isolated and enriched prior to the MS analysis, an approach chosen by most of the studies dealing with the investigation of the lysosomal proteome. Unlike other sample preparation strategies employed in combination with untargeted DDA-MS studies such as peptide fractionation [83], lysosome enrichment strives for the isolation and preservation of the intact lysosome. Enrichment of intact lysosomes can be accomplished either within the course of a broader subcellular fractionation process, where the lysosome-enriched fraction (LEF) is one fraction among several other fractions of interest, or by employing a protocol specifically designed for the enrichment of lysosomes [9]. Various protocols for lysosome enrichment have been developed. Shared among these protocols is the initial, deliberate disruption of the plasma membrane. The recovered the post nuclear supernatant (liquid portion of the cellular lysate after removal of the cell nuclei, intact cells and membrane fragments) is then further used in the enrichment procedure. Based on the enrichment principles and techniques used, three approaches can be distinguished.

With density gradient centrifugation-based approaches, the recovered post nuclear supernatant is further separated by ultracentrifugation into individual organelle-containing fractions [9]. This approach takes advantage of organelle-specific densities and its thereby resulting sedimentation behaviour in a density matrix such as sucrose [109]. For a better separation of organelles i.e. reduced overlap of organelle fractions, numerous density gradient matrices other than sucrose, which is characterized by overlapping sedimentation between mitochondria and the lysosome fraction have been introduced [110]. Similar to this idea is the method to actively manipulate lysosomal density by loading hepatic lysosomes of rats via the endocytic pathway using non-ionic Triton WR-1339 [111]. This results in a density shift of these lysosomes, called tritosomes, which enhances their discrimination from other organelles by ultracentrifugation.

Approaches employing immunoprecipitation (IP) make use of immobilized antibodies or streptavidin beads. For the capture of intact lysosomes, a lysosomal surface protein like lysosome-associated membrane protein 1 (LAMP1) or transmembrane protein 192 (TMEM 192), is fused to an epitope tag such as 2x/3x FLAG [25, 112], 3x HA [113] or a 2x Strep tag [114]. These fusion proteins are then stably expressed in the cell of interest. Lysosomes are selectively bound by tag directed antibodies and the intact lysosome can be separated via the membrane-anchored target protein binding from the remaining post-nuclear supernatant.

Superparamagnetic iron oxide nanoparticles (SPIONS) allow for the enrichment of lysosomes from cells in cell culture utilizing coated, submicron particles [115]. Following the uptake of SPIONS by the cells through unspecific fluid phase endocytosis and the subsequent accumulation of particles within the lysosomes, intact lysosomes can be retained from the cell homogenate by a strong magnetic field [56, 58].

The success of a lysosome enrichment is typically evaluated in a follow up experiment. To assess the integrity of eluted lysosomes, an enzyme assay specific for the detection of the lysosomal luminal protein β hexosaminidase is employed [116]. Antibodies directed against organelle-specific marker proteins, in combination with Western blot detection allow an assessment of enrichment efficiency. Each enrichment technique has its advantages and comes with certain limitations. While density gradient-based approaches can be applied to samples of any type, ranging from cell culture to tissue samples, recovery rates are usually lower and the obtained lysosomal fraction is more prone to impurities from other organelle fractions [110]. The enrichment with SPIONS provides high yields and excellent enrichment [117]. Its application however is limited to cell lines, and here exclusively to those cell lines with a functional endocytic transport system thereby excluding cell lines with an impaired endo-lysosomal transport as often seen in disease and disease-related conditions [118]. Moreover, the utilization of SPIONS requires the determination of cell line-individual culture and pulse-chase conditions [56]. Immunoprecipitation or -capturing of intact lysosomes represent the fastest technique to enrich for lysosomes. However, purified lysosomes cannot be released intact; only lysosomal proteins after a lysosome rupture can be recovered. Despite the speed and simplicity of this approach, it's important to consider the upfront work involved in this strategy, which includes the generation of the target fusion protein construct and stable cell line, as well as the establishment of IP conditions [25, 112]. Singh et al. [117] evaluated the three different approaches (SPIONs, sucrose-density gradient and TMEM192-IP) for the generation of enriched lysosomes applying DIA mass spectrometry. A notable observation was, MS after sucrose density gradient centrifugation identified only 55% of the currently confirmed 341 lysosomal proteins. Both the SPION and TMEM approaches demonstrated an improved lysosomal proteome coverage rate (71% for each method), however the study highlights that a considerable portion of the lysosomal proteome remains undetectable and thereby unexplored by currently available lysosome enrichment strategies [117]. Considering the individual advantages and limitations associated with the techniques utilized for lysosome enrichment, it becomes evident that there is no universally applicable enrichment method suitable for cell lines and tissues. Additionally, a common limitation with all enrichment strategies is the omission of patient sample material from MS analysis, since those samples are typically available in only limited quantities [119].

This lack of a “one-size-fits-all” approach prevents a systematic investigation of the lysosomal proteome across cell lines and tissues under reproducible conditions. An approach,

circumventing the previously discussed drawbacks and challenges associated with enrichment and standard MS approaches, can be the implementation of targeted MS strategies.

1.6. Targeted MS for quantification of low abundant proteins from complex samples

Targeted MS proteomics encompasses user-defined data acquisition for selected proteins rather than a comprehensive examination of the entire proteome. For this purpose, targeted data acquisition (and follow up data analysis) focuses on a set of pre-defined peptides which are proteotypic, i.e. unique to the protein(s) of interest, operating the MS instrument in a selective data acquisition mode [120, 121]. In the course of this work two approaches namely Multiple Reaction Monitoring (MRM) and Parallel Reaction Monitoring (PRM) were used for the targeted quantification of lysosomal proteins from whole-cell and tissue lysates.

MRM experiments are typically performed on triple quadrupole instruments, an MS instrument configuration where three quadrupole mass analyzers are arranged sequentially [122] and quadrupoles one (Q1) and three (Q3) are used as mass filters (Figure 6). The first quadrupole is set to a narrow m/z range, selectively isolating the pre-defined intact peptide precursor of interest. Following the fragmentation of the isolated, intact peptide ion in quadrupole two (Q2), m/z windows are set which allow only pre-defined product ions to pass sequentially quadrupole three (Q3). Product ions are detected by a secondary electron multiplier [122] providing an excellent linear range and sensitivity [73, 91]. Usually three to five precursor-product ion m/z combinations, so-called transitions are monitored per peptide. The multiplexing of this two-step filtering strategy makes MRM highly specific for the peptide(s) of interest and reduces the interference from unrelated compounds in the samples [120]. As a result, even the detection of proteins in the low attomole range from complex samples [123], the determination of protein copy numbers lower than 50 copies per cell [124], and the reproducible quantification of lowest protein amounts from body fluids (in the low picogram per milliliter range) such as human plasma/serum is possible [125]. Unlike MRM experiments, which are based on precursor-product ion m/z combinations, PRM experiments require only the definition of a precursor m/z [126]. This is due to the different configuration of mass analyzers in the instrument setup. PRM experiments are performed on Q-OT [126] or Q-TOF instruments [127]. In this instrument configuration a low-resolution quadrupole mass analyzer and the high-resolution mass analyzers are combined in one hybrid instrument. The quadrupole acts as a mass filter isolating the pre-defined intact peptide ion, the high resolution mass analyzer enables the acquisition of a full MS2 spectrum at high mass accuracies [120] (Figure 6). The simplified implementation of a PRM experiment makes it an attractive alternative to traditional MRM experiments providing greater flexibility in selecting the most suitable fragment ions (post-data acquisition fragment ion selection) and the high resolution, full MS2 spectra results in an increased selectivity as compared to triple quadrupole instrument-based MRM experiments [120, 128].

Multiple studies directly comparing both approaches (PRM and MRM) demonstrated a closely comparable quantitative performance, revealing that discovered disparities and superiorities being of non-systematic, marginal nature [128-130].

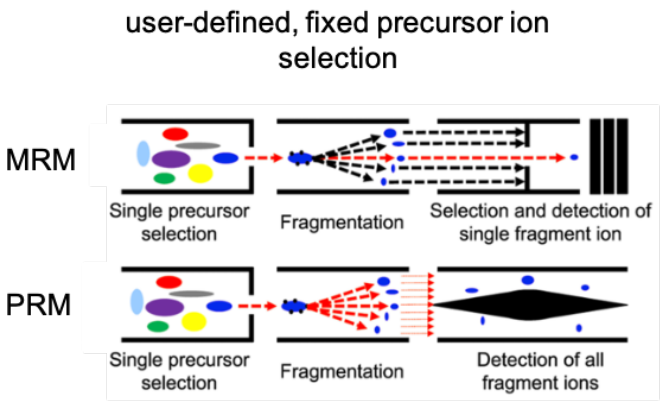


Figure 6: Operational mode of an MS instrument for targeted data acquisition. In MRM and PRM mode, the intact precursor ion of interest, defined by the user, is isolated and fragmented. A full fragment ion spectrum is recorded in PRM-MS, while in MRM-MS only (pre-)selected fragment ions are detected and recorded. Abbrev.: MRM: multiple reaction monitoring; PRM: parallel reaction monitoring. Figure from [92].

Key parameters in a targeted MS experiment which have to be controlled, include the so-called dwell, maximal ion injection, and cycle times. In an MRM experiment, the dwell time specifies the duration of the triple quadrupole instrument remaining on one transition setting (precursor/fragment ion combination) for signal acquisition before switching to the next transition [131]. In PRM data acquisition, where no single transitions are recorded, the maximal ion injection time which is the maximum time the precursor is isolated and prepared for MS2 spectrum readout, can be set [126]. Increasing dwell or ion maximal injection times of the targets enhances sensitivity. However, with an increasing number of peptide targets to measure, also the cycle time of the assay (time needed to iterate once through the complete list of defined targets) increases. In return this reduces the number of data points per peak and can compromise the quantification accuracy of the peak [131]. To alleviate this shortcoming, scheduled data acquisition strategies were devised. Instead of monitoring all peptide targets in a PRM experiment, respectively the selected MRM-transitions continuously, in a dynamic or scheduled experiment the peptide targets are monitored only around their projected or determined time of elution [132]. As a result, sensitivity for the precursors or transitions monitored can be maximized at the time of detection, while the number of targets in an MS experiment can be expanded making targeted MS a highly efficient approach for data acquisition. The strength of targeted MS data acquisition approaches is based on its independence of precursor abundance, which guarantees a reproducible data acquisition of even low abundant peptides from complex samples at highest sensitivity [121, 133] and quantitative precision compared to the conventional DDA approach. However, employing targeted MS approaches for data acquisition first requires an upfront phase of assay development and optimization before the actual MS experiment can be performed. This initial phase of assay development will be discussed in the next section.

1.6.1 Targeted MS assay development and optimization

In contrast to unbiased and untargeted DDA and DIA approaches that can be readily applied, targeted MS approaches focus on a list of predefined, curated peptide targets. The selection, validation and optimization of peptide targets is typically conducted manually or semi-automatically, involving multiple steps that require a significant amount of work. Over the years a vast body of literature providing guidance on this topic has been generated [131, 134-137]. The following section will summarize and explain the basic principles and steps involved in targeted MS assay development. The starting point in targeted MS assay development is the screening and selection of suitable peptides from the protein(s) of interest. Although proteolytic digests can result in hundreds of peptides per protein, not all of them are equally well detectable [138]. Therefore, filtering and evaluation of the peptides, representative for the protein of interest is required to select only the most favorable peptide(s). These very few suitable peptides selected for the final MS experiment also known as proteotypic peptides, stand out due to their specific properties and behaviour in the MS experiment. An indispensable criterion involves the necessity of peptide uniqueness, where the peptide's sequence should be exclusively matched to the protein of interest or its protein isoform. Additionally, the selected peptide should show a strong MS response [138, 139]. Further rules and recommendations guide the selection of surrogate peptides, encompassing considerations related to their amino acid composition. These guidelines aim to prevent for example peptides containing amino acids prone to (non)-specific modifications, post-translational modifications (PTMs), or missed cleavages, since this might adversely impact the precision and accuracy of quantification [140].

In practice for peptide (pre)-selection, researchers often rely on previously acquired MS data, such as from a DDA or DIA experiment. Alternatively, one can rely on public MS data repositories [140-144] and associated selection tools [145, 146], which prove to be beneficial in generating an initial peptide target list for PRM experiments or a list of the most responsive transitions for designing an MRM assay. Since experimental conditions can vary between studies, peptides identified as potentially representative are re-assessed for their detectability with the specific experimental setup and LC-MS instrument configuration designated for the later assay. In this phase of method refinement, multiple precursor charge states per peptide and a larger number of transitions (usually exceeding five transitions per peptide charge state for MRM assay development) are typically evaluated. Ultimately, only the peptide precursor respectively the three to five MRM transitions yielding the highest signal are selected for the final quantitative assay, thereby maximizing measurement sensitivity (Figure 7). To ensure the identity of peptides, a peptide standard covering the targeted peptides of interest can be applied. Such a peptide standards facilitate signal extraction and confirmation of the optimal MRM-transitions [147], as well as the determination of peptide-specific retention times. However, since peptide retention times depend on variable conditions such as LC gradient length and design, or sample type and complexity, the concept of dimensionless,

indexed retention time (iRT) to normalize peptide retention time was applied in this work [148]. Retention time indexing implies that each target peptide's absolute retention time is assigned a numeric, indexed value on a standardized scale (iRT-score) which is predetermined by a set of unique reference peptides (iRT-peptides). In addition, during MRM assay design, MS-instrument-specific full fragment ion spectra (MS2 spectra) can be acquired using the MRM-initiated detection and sequencing acquisition (MIDAS) approach [149]. In the following stages of assay development, chosen targets i.e. peptide precursors (in the case of a PRM experiment) and transitions (in the case of a MRM experiment) are validated and, if necessary, modified and optimized. For that reason, the MS response of selected targets is further tested in a complex sample background, re-assessing its quantitative performance and suitability under real assay conditions. Unspecific signals should be minimized and co-eluting peaks avoided as far as possible [131]. Optionally, also applied collision energies (CE) for peptide fragmentation can be optimized to boost the signal response for single peptides [150]. In case stable isotope-labeled peptide standards are employed for quantification, the spike-in amount of the peptide standard relative to the endogenous peptide signal, has to be adjusted. At the end of the assay development process, the targeted MS assay is finalized (Figure 7). MS2 spectra of peptide targets acquired during the development process are compiled in a spectral library, serving as a reference for manual or automated, software-based peak detection and evaluation [151].

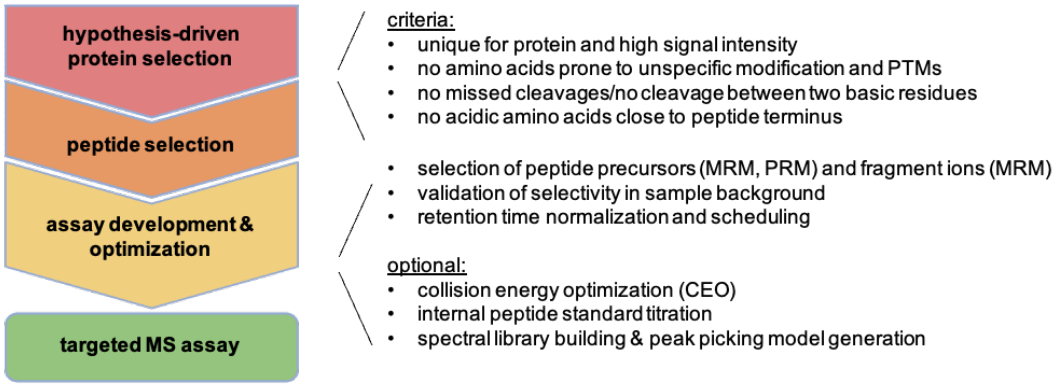


Figure 7: Workflow for targeted MS assay development and optimization. Targeted MS assay development and optimization is a multi-step procedure. Shown are the different phases as well as their associated criteria or tasks. Abbrev.: CEO: collision energy optimization; MRM: multiple reaction monitoring; PRM: parallel reaction monitoring; PTM: post translational modification.

Acquired iRT-scores and their derived absolute peptide retention times act, beside the definition of precursor m/z for PRM respectively the MRM transitions, as another discriminator of peptide identity [148]. Furthermore, as the iRT-score reflects the relative elution order/retention time of target peptides compared to the calibrant peptide(s) in an assay, calibrant peptide retention times allow prediction of absolute retention times for the peptides of interest in a planned targeted MS experiment

[148]. Like this, iRT-scores facilitate scheduled data acquisition and ensure the transferability of a developed assay across different LC-MS configurations as well as sample types.

1.6.2 Application of targeted MS for the investigation of lysosomal proteins

Targeted MS techniques like Parallel Reaction Monitoring (PRM) and Multiple Reaction Monitoring (MRM) are still relatively uncommon for the analysis of lysosomal proteins with experiments mostly conducted using clinical/patient samples. Clinical samples include urine, cerebrospinal fluid, dried blood spots, and occasionally tissue or cancer samples [119]. The (endo-) lysosomal peptides monitored in these experiments were investigated in the context of studying neurodegenerative diseases such as Alzheimer's (AD), Huntington's disease (HD) and Parkinson's (PD), where lysosomal proteins could serve as potential biomarkers [152-155]. Notably, only a few clinical studies that directly screen for LSDs exist [156-159]. Nevertheless, given that these clinical investigations only focus on a limited number of peptide targets (usually fewer than five lysosomal proteins per assay) such as cathepsins [160] and lysosome-associated membrane proteins [161], a broader view on the lysosomal proteome remains unattainable.

1.7 Antibody-and aptamer-based microarrays for the study of lysosomal proteins

Alternative techniques for the targeted study of proteins not relying on MS are immunoaffinity-based and aptamer-based microarrays, using the OLink and/or SomaScan platform. While aptamer-based microarrays (e.g. SomaScan) make use of specific protein-nucleic acid interactions, directly binding the target protein to a short, single stranded oligonucleotide (aptamer) [162, 163], immunoaffinity-based microarrays (e.g. OLink) apply pairs of antibodies with each antibody coupled to a single stranded complementary DNA sequence hybridizing upon specific protein-antibody pair interaction [164, 165]. In both approaches the unique DNA sequences used act as barcode identifying the detected antigen-antibody pair or antigen-nucleic acid interaction. The DNA sequences quantitatively convert the amount of protein per sample into a nucleotide-based signal for quantification [163] that can be further amplified by PCR. The technique's dependence on an antigen-antibody pair or antigen-nucleic acid interaction shows strengths and weaknesses. Depending on the quantitative performance of the targeted MS assay, immunoaffinity-based and aptamer-based microarrays can excel in sensitivity, dynamic range, and sample throughput [163, 165]. However, the biomolecular affinity interaction also limits the technology's applicability and flexibility. Only predefined protein targets with biomolecular interaction partner, as mapped and validated by the assay vendors and contract research partners are available. This excludes any customization of the assay such as the addition of a new protein target, or modification in case the study of PTMs or protein isoforms is planned. Currently, the OLink and SomaScan platform cover 87 (25%) [166] and 173 (50%) [167] lysosomal proteins for quantification respectively, as compared to all known lysosomal proteins.

1.8 Strategies in MS-based quantitative proteomics

To study a complex biological system using MS, both the reliable detection, but also the precise and accurate quantification of proteins within and across multiple samples is necessary. Similar to the protein identification, where peptide identifications allow the inference of a protein, available peptide quantifications associated with the same protein per sample/condition are aggregated into the protein quantification [168]. Peptide quantification can be performed at either the MS1 level (intact peptide/precursor ion) or at the MS2 level, considering the fragment ions of the precursor ion [120, 169]. In either case, peptide-specific MS signals must be extracted over time from the recorded data (Figure 8). Additionally, the MS signal intensity at the apex of the individual chromatographic elution profile (peak) or its area can be used for quantification [170].

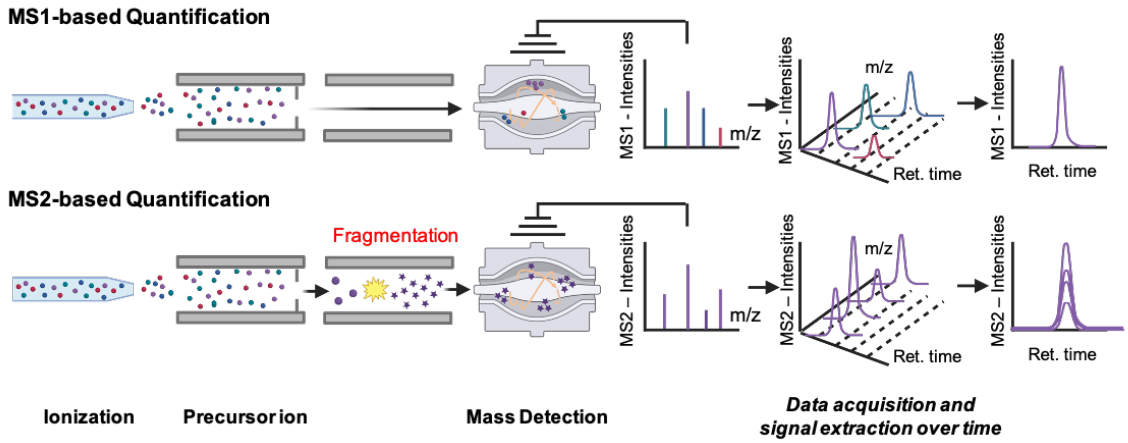


Figure 8: Schematic view on MS1- and MS2-level based quantification. Shown is the operational principle of the MS instrument, here a Q-OT hybrid instrument, acquiring either MS1 signals (intact precursor intensities, upper panel) or MS2-signals (fragment ion intensities, lower panel) in DDA mode. Precursor and/or fragment ion intensities acquired throughout the chromatographic separation are extracted and MS signals are (re-)constructed for quantification. Abbrev.: DDA: data-dependent acquisition; Q-OT: quadrupole-Orbitrap. Figure was created with BioRender. Figure is based and adapted from [91].

Quantitative MS is accompanied by a variety of sample preparation- and MS-related challenges. One of these are individual ionization efficiencies for different peptides [74], which renders MS not inherently quantitative for protein quantification. As each peptide's individual amino acid combination "sets" the peptide ionization efficiency, equal amounts of different peptides do not necessarily result in the same MS response, i.e. signal intensity. Consequently, only the MS signal of the same peptide/protein across two or multiple samples can be compared, resulting in a ratio or relative change in protein abundance for the compared conditions which is known as relative quantification. Further, also the sample matrix and its complexity affect the ionization of analytes, influencing the signal intensity of the same peptide across samples of varying complexity and types, a phenomenon known as sample matrix effect [171, 172]. Therefore, direct relative quantification is only possible

between identical or similar sample types. Absolute quantification strategies, on the other hand, rely on the spike-in of externally quantified standard peptides and enable a direct comparison of peptide/protein abundances within and across samples and sample types. As the name implies, with this approach absolute protein quantities can be determined [74, 168]. In the following chapters, the most common label-free, label-based and internal standard-based strategies for relative and absolute quantification are reviewed (Figure 9).

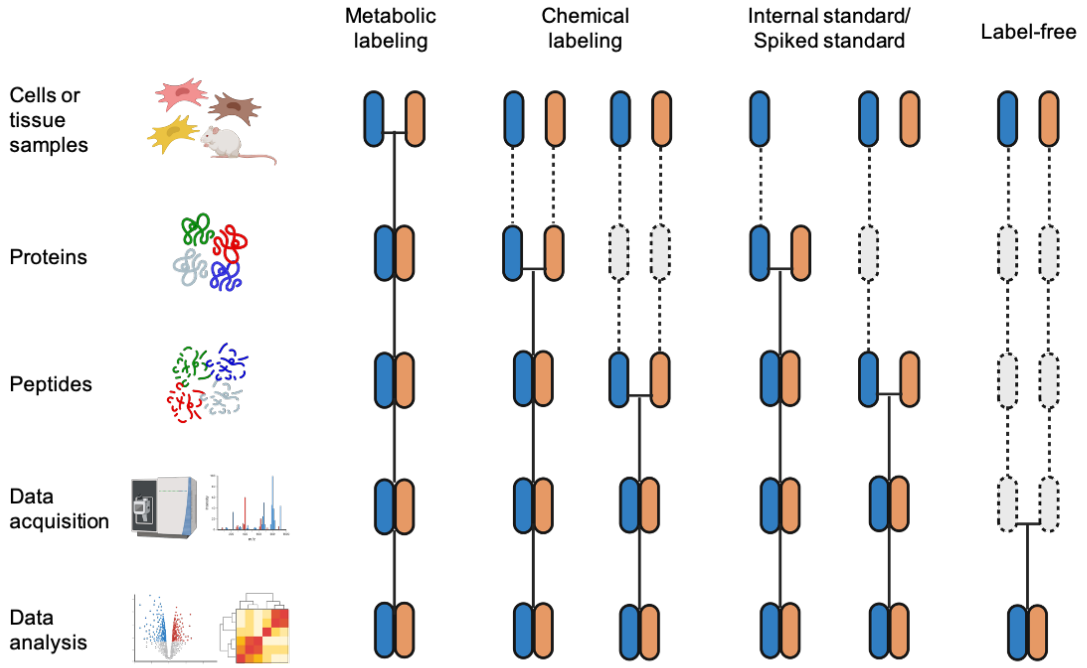


Figure 9: Overview of the most common workflows applied in quantitative MS-based proteomics. The scheme shows major stages of a quantitative MS-based experiment. To be distinguished are label-based approaches, such as metabolic or chemical labeling; label-free and internal standard-based quantification approaches. Colored boxes represent (two) different experimental conditions. Horizontal lines indicate the stage when samples are combined. Dashed lines highlight stages at which experimental variation may arise, contributing to the overall quantification error. Figure is based on [168].

1.8.1 Label-free quantification

Label-free quantification (LFQ) is the most straightforward and economical approach in relative quantitative proteomics. Avoiding any form of sample labeling, or the use of spiked, internal standard peptides, the approach is based on the linearity of the MS signal response for peptides across approximately four orders of magnitude, correlating with their peptide concentration [173]. The conceptual simplicity extends to its workflow where each sample is individually prepared and MS data is acquired separately. While this simplicity promotes a maximum degree of flexibility, and allows LFQ from all sample types and specimen in studies with theoretically unlimited sample size,

the simplicity of LFQ brings along challenges. Due to separate processing of samples, LFQ is more prone to variability throughout the proteomics workflow, resulting in a higher quantitative variance (Figure 9). Therefore, a standardized and reproducible sample preparation, as well as a robust LC-MS setup for data acquisition, is required to minimize variability [168].

For quantification, extracted ion chromatograms (XICs), typically derived from precursor ions (MS1) as seen in DDA, or either of the fragment ions (MS2) as seen in DIA, are employed [174]. The comparison of these extracted signals between the different samples, which is essential for the actual relative quantification, introduces an additional challenge to LFQ. Since data for different samples was recorded individually, during post-acquisition phase, the extracted peptide signals of the same peptide species must be mapped and aligned across multiple LC-MS measurements [168]. Even though label-free experiments cannot provide absolute protein quantities, which are typically required for a direct comparison of abundances between different proteins, multiple approaches for label-free absolute quantification estimation, further referred to as label-free absolute quantification have been developed. Key to label-free absolute quantification is the observation that protein abundance correlates with the number of identified peptides and/or peptide spectral matches (PSM) [175]. While initially, label-free absolute quantification approaches counted the number of peptides or PSMs per identified protein [168], modern label-free absolute quantification strategies such as intensity-based absolute quantification (iBAQ) use the sum of all peptide intensities per protein. To normalize for a protein size bias (larger proteins generate more peptides than smaller proteins), values are divided by the number of theoretically detectable peptides [176]. Computed iBAQ values enable a rough approximation of protein abundances, allowing for comparisons across different proteins. However, it should be stressed that label-free absolute quantification is generally less accurate than label-based absolute quantification approaches [168] and cannot provide unit-based protein quantities.

1.8.2 Label-based quantification

Label based quantification strategies apply non-radioactive, stable isotopes, typically ^{13}C, ^{15}N and ^{2}H to label different proteomic conditions (samples). While physicochemical properties of the analyzed labeled peptides, such as retention time and ionization efficiency remain largely unchanged upon stable isotope labeling (exceptions are deuterated amino acids resulting in a retention time shift [177]), differentially stable isotope-labeled peptide ions and/or their fragmentation products become distinguishable by MS due to their mass difference. This allows labeled samples to be combined during the proteomics workflow, reducing variability and facilitating relative quantification (Figure 9). Therefore, MS signals across the different isotope-encoded states (distinct sample conditions) determined from the same MS experiment (same sample) are directly compared. Labeling with stable isotopes can be accomplished either through metabolic, chemical [168] or enzymatic [178] labeling strategies. Since the field predominantly applies metabolic and chemical labeling strategies

for quantification experiments, the upcoming section will exclusively cover these two labeling strategies.

1.8.2.1 Metabolic labeling

In metabolic labeling, stable isotopes are introduced into living organisms (*in vivo* labeling) through their nutrient source, e.g., amino acids and salts [179]. As the cells undergo growth and division progressing through the normal cell cycle, cells gradually incorporate the provided stable isotopes, replacing the natural isotopes in newly synthesized proteins [180]. The stable isotope labeling by amino acid in cell culture (SILAC) technique is a widely used metabolic labeling approach. In a classic SILAC experiment, two (or more) populations of cells [181, 182], representing, for example, different experimental conditions, are cultured in individual media. Each medium is supplemented with isotopically distinct forms of amino acids, utilizing either natural isotopes (non-labeled, light standard medium) or distinct combinations of heavy isotopes ^{13}C, ^{15}N and ^{2}H for amino acid labeling. When conducting a tryptic digestion experiment these are typically arginine (R) and lysine (K), although initially leucine (L) was utilized [180]. When a label incorporation of more than 95% is reached [180, 183], i.e. an (almost) complete replacement of natural isotope containing amino acids with the supplemented stable isotope-labeled amino acid analogs is achieved, cells are harvested and lysed. For the MS experiment equal amounts of cell lysate are pooled (Figure 9), and further subjected to a combined sample processing followed by MS data acquisition [180]. Quantification is based either on the MS1 signal in case of an untargeted MS experiment, or in case of a targeted MS experiment on the MS2 signal. In both cases, the MS signals are compared between the unlabeled and isotope-labeled intact peptide ions (MS1-based quantification) or fragment ions (MS2-based quantification), showing the distinct mass shift introduced by the metabolic incorporation of stable isotope-labeled amino acids.

Metabolic labeling strategies like SILAC stand out due to their immediate combination of samples (experimental conditions) at the earliest possible timepoint. Combined sample processing and data acquisition reduces quantification variability significantly, leading to a higher quantification accuracy and precision [168]. Initially the SILAC technique was designed as a mammalian cell culture experiment, however its labeling principle was successfully transferred to bacteria such as *E. coli* [184], as well as yeast [185], plants [186] and other model organisms [181]. Even the labeling of higher organisms, such as zebrafish [187] or mice [183], has been demonstrated. One major application of the SILAC technique lies in expression proteomics, but others exist [181]. Expression proteomics which aims to characterize protein changes among various sample conditions using the SILAC technique as described, however falls short in reflecting the dynamics of protein expression in an organism. As a consequence, multiple variants of the SILAC technique have been developed [168], so called pulse-SILAC methods [176], and are used today in studying the turnover of individual proteins. Another metabolic labeling strategy, ^{15}N metabolic labeling, has been applied to label

multicellular organisms such as *C. elegans* (nematode), *D. melanogaster* (fruit fly) [188] and rat [189]. In this method, the multicellular organisms are fed with ^{15}N-enriched food (*E. coli*, yeast and algae) for label incorporation. Unlike the SILAC approach, which specifically labels only arginine and lysine by introducing a fixed number of stable isotopes to the amino acid structure, ^{15}N metabolic labeling replaces each natural nitrogen isotope (^{14}N) with the ^{15}N stable isotope counterpart. As a result, the number of ^{15}N labels varies among different peptides and the ubiquitous presence of ^{15}N in the peptide leads to broad, complex isotope patterns which complicates data analysis [168].

1.8.2.2 Chemical labeling

Chemical labeling of proteomic samples involves the use of chemically reactive, mass tags. Mass tags are covalently attached to proteins or peptides through their reactive groups after protein biosynthesis (*ex vivo*) (Figure 9). Various chemical tagging approaches have been developed. One can distinguish between isobaric labeling and differential mass labeling (non-isobaric) methods. Isobaric labeling involves techniques like e.g. tandem mass tags (TMT) [190] and isobaric tag for relative and absolute quantitation (iTRAQ) [191]. Differential mass labeling comprises methods such as isotope-coded affinity tags (ICAT) [192] or mass differential tags for relative and absolute quantification (mTRAQ) [193] as well as dimethyl labeling [194]. The two most popular chemical labeling approaches TMT and iTRAQ are briefly introduced in the following section.

Isobaric labeling strategies such as TMT and iTRAQ employ a set of multiple isobaric mass tags to label different samples. Although TMT and iTRAQ reagents (mass tags) can be distinguished structurally by their reporter and balancer regions, both techniques rely on the same labeling and quantification principle. The N-hydroxysuccinimide ester-based reactive group of the mass tag targets the N-terminus of peptides and the side chain of the amino acid lysine (both primary amines). The isobaric mass tag regions share the same nominal mass, but show, depending on the isobaric mass tag “variant”, a differential distribution of stable isotopes within the reporter and balancer region. This results in reporter ion masses ranging from 126 to 134 m/z for TMT [195], and 114 to 117 m/z for iTRAQ [196], respectively. In an MS experiment, the distinct isobaric mass tags appear when derivatized to the same peptide (e.g. originating from different labeled samples) as one single signal/peak in the precursor ion scan thereby remaining indistinguishable at the MS1 level. Following precursor isolation and fragmentation, the sample-specific reporter ions are released and can be observed in the lower m/z range of the fragment ion spectrum (MS2). Here, the ratio of the reporter ion intensities reflects the relative abundance of the peptide in the different samples [190]. Chemical labeling strategies have the unique advantage over metabolic labeling approaches, that they can be applied to any type of biological sample, including human tissues, body fluids etc. [197]. This flexibility stands out when compared to metabolic labeling, which is limited to living organisms. Furthermore, isobaric labeling strategies excel through its multiplexing capabilities, allowing up to 18-plexes (samples) in a TMT experiment to be combined in one single experiment [195]. In

combination with peptide fractionation and/or peptide enrichment, isobaric mass tagging facilitates deep profiling studies, allowing for the systematic and reproducible quantification of low(er)-abundance post-translational modifications (PTMs) [198, 199].

1.8.3 Internal standard-based quantification

Internal standard-based quantification involves the consistent addition of a defined amount of an externally generated standard to each sample prior to MS analysis. Often referred to as a spike-in experiment, the added internal standard serves as a reference for quantification (Figure 9). Depending on the timepoint of introduction, it accounts for variations either in MS data acquisition or in both sample preparation/handling and MS data acquisition [168]. In general, one has to distinguish between peptide isomers that are sequentially derived from their parent peptide, i.e. from the peptide of interest [200, 201] and naturally occurring standards which are sequence-identical to endogenous peptides [168] (Figure 10).

Since naturally occurring standards (peptides) would be indistinguishable from the sample upon spike-in, such an internal standard has to be stable isotope-labeled using heavy amino acids [202]. The ratio of a signal intensity between the co-eluting, (heavy) labeled internal standard (peptide) and the native peptide (non-labeled) reflects the relative abundance and can be compared across different samples and experimental conditions (Figure 10).

An interesting alternative to stable isotope-labeled, sequence-identical standard peptides that are often associated with high costs, is the use of peptide isomers as standards. Such a peptide isomer-based standard is a synthetic i.e. non-endogenous peptide construct, which has an identical amino acid composition as the peptide of interest but its structural arrangement of amino acids differs [200] (Figure 10). Although indistinguishable based solely on mass at the MS1 level, the positional permutation of two adjacent amino acids at the N-terminus of the peptide isomers renders them distinguishable by their sequence-specific fragmentation pattern (MS2 level) and distinct retention time. Yet, the separated peptide isomers called minimally permutated peptide analogs (MIPAs) are assumed to exhibit practically the same ionization efficiency, allowing their use as a stable isotope-labeled-free isobaric internal standard. Employing a quartet of structural peptide isomers (of identical mass and similar/same ionization efficiency), which comprises two subsets of N-terminally permutated dipeptide motifs further enables a multiplexed internal standard-based quantification. While in theory peptide isomer-based standards can be generated for every peptide of interest, certain recommendations apply for the selection of peptides and amino acid permutation, ensuring e.g. unambiguous fragment ion-based isomer identification [201].

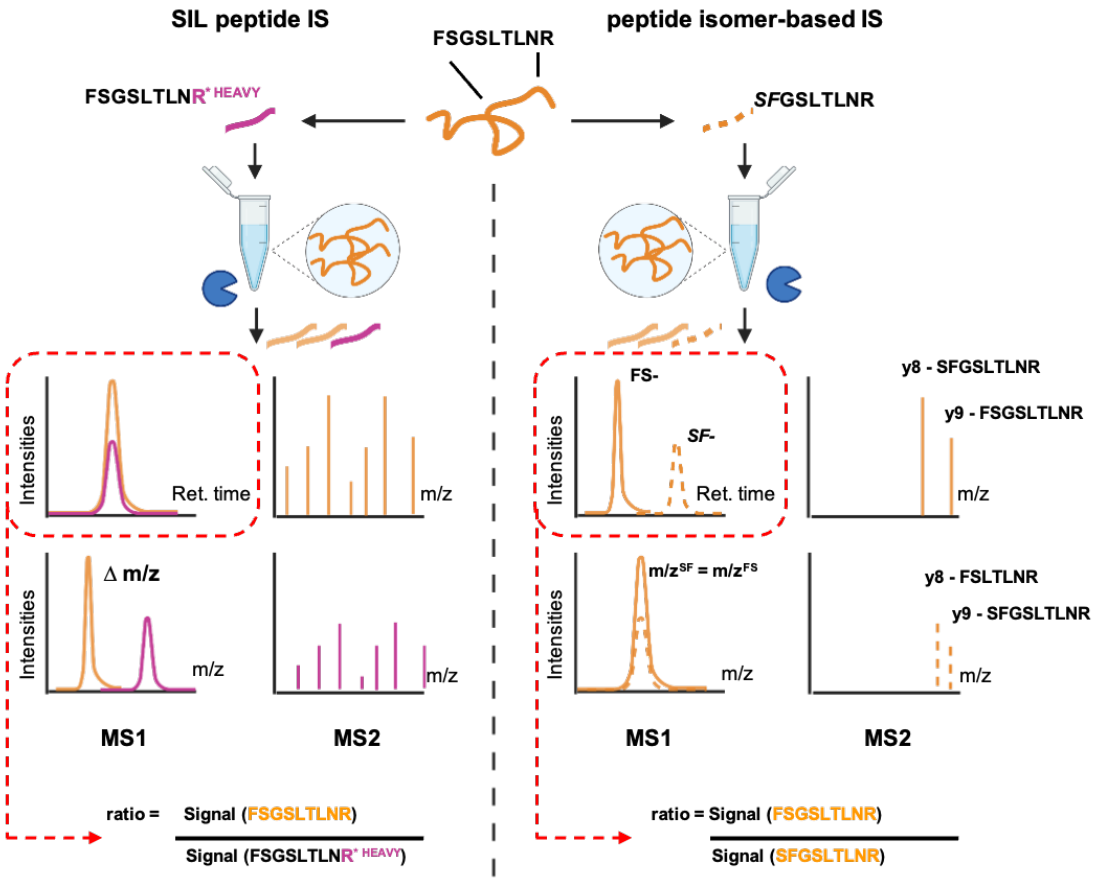


Figure 10: Principle(s) of internal standard peptide-based quantification. Shown are two strategies utilizing either sequence-identical, stable isotope-labeled (SIL) peptides (left) or peptide isomers with positional permutation (right). In both cases, the internal standard is added to the sample. Depending on the standard type, the endogenous peptide and internal standard are distinguishable by peptide retention time (peptide isomer), MS1- and/or MS2-mass(es) (peptide isomers and sequence-identical, SIL peptides). An identical (SIL, sequence-identical peptides) or very similar ionization efficiency (peptide isomers) allows for relative and absolute quantification, with the internal standard peptide signal serving as reference. Abbrev.: IS: internal standard; SIL: stable isotope-labeled. Figure was created with BioRender. Figure is based and adapted from [140] and [201].

If known quantities of either a stable isotope-labeled reference standard (peptide) or a peptide isomer standard are added to the sample, not only relative quantities but also absolute peptide amounts, and by inference protein amounts [168], can be determined. Therefore, the MS signal of the internal standard with known standard amount is compared with the MS signal of the sample [140, 203]. Absolute protein quantities are of great value. Other than relative quantities, absolute protein quantities enable the study of protein complexes and the discovery of protein stoichiometries [204, 205]. In systems biology, absolute quantification information is needed for the generation of mathematical models simulating complex biological systems [206, 207]. In the clinical context, accurate absolute quantification of proteins is essential for the identification and validation of potential biomarkers, thereby establishing reference values for disease-relevant proteins [208, 209].

1.9 Generation of internal standards for MS based absolute quantification

Reference standards for absolute quantification can be generated either as peptides, as exemplified in the absolute quantification (AQUA) [203], quantification concatemer (QconCAT) [206], and Protein Epitope Signature Tag (PrEST) [210-212] methods, or as full-length proteins using the absolute SILAC [213], Full-Length Expressed Stable Isotope-labeled Proteins for Quantification (FLEXIQuant) [214] or Protein Standard Absolute Quantification (PSAQ) [215] strategy. Internal standards used throughout my work were derived either through the AQUA method or the QconCAT strategy. Therefore, the focus of this introductory chapter will be on these two methods, nevertheless the other methods will be briefly outlined and put into context.

The AQUA strategy [215] is the most straightforward approach and involves the chemical synthesis of stable isotope-labeled peptides which are sequence-identical to proteotypic peptides of interest. The quantification of the generated synthetic, stable isotope-labeled peptides is achieved through amino acid analysis [216] and defined amounts of quantified peptide standards are added to the biological sample either during or after protein digestion. Often referred to as gold-standard for absolute quantification, AQUA peptide standards carrying a PTM enable the study of PTM stoichiometries [217]. However, due to the high cost of synthetic, absolutely quantified AQUA peptides, the AQUA strategy might not be the most suitable option when hundreds or thousands of absolutely quantified peptide standards are required, performing absolute quantification on a large(r) scale. One efficient way to generate a large number of peptides without relying on chemical peptide synthesis is the QconCAT technology [206] (Figure 11). Here, utilizing recombinant DNA techniques, artificial QconCAT genes are designed which encode for artificial proteins that concatenate proteotypic peptides of interest, so called Q-peptides. The synthetic QconCAT protein constructs are expressed in an *E. coli* host strain [206], however also cell-free expression is possible with the MEERCAT technology [218], supplementing the cell media with stable isotope-labeled amino acids for metabolic labeling. His-tag purified and quantified, heavy labeled QconCAT proteins can either be co-digested with the biological sample, or digested separately and then spiked into the biological sample. In both scenarios stable isotope-labeled Q-peptides are being released in an equimolar stoichiometry and function as an internal standard for absolute quantification of unlabeled, endogenous sample peptides. With the QconCAT strategy, up to 25 to 30 peptides can be combined and generated from one QconCAT protein. The *in silico* design of constructs enables the tailored assembly of signature peptides (Q-peptides) which can originate from various proteins, but are combined finally in one construct [206]. Using this approach, an extensive quantitative study demonstrated the absolute quantification of approximately 2000 proteins in yeast [219]. Unlike AQUA peptides, the QconCAT standard can undergo active proteolytic digestion with the biological sample. This allows to account for sample preparation-related systematic errors, as both the sample and the internal standard are equally affected by variations and sample losses. However, thorough attention

needs to be given to the extent of QconCAT's proteolysis [220]. Due to the concatenation of peptides in the artificial protein construct (QconCAT) the sequence context of the individual QconCAT peptides, i.e. the adjacent peptide sequences, does not resemble the natural peptide sequence context. This discrepancy might lead to an altered cleavage behavior/efficiency, causing differences in peptide release (compared to the natural, endogenous peptides) which could potentially impact the accuracy of quantification [206].

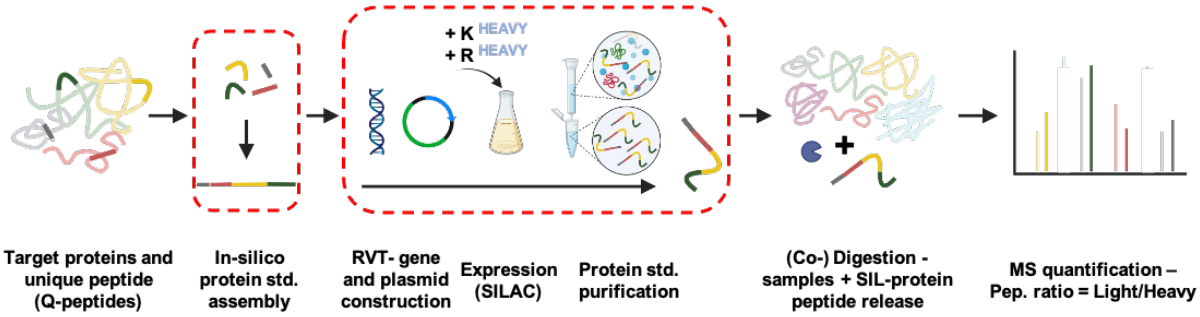


Figure 11: Concept of QconCAT-based protein quantification. Shown are the main stages necessary for a QconCAT-based protein quantification. Representative peptides are selected for the proteins of interest and concatenated into an artificial protein construct/standard. Following reverse translation (RVT), the synthetic gene product introduced into an expression vector is expressed, metabolically labeled using stable isotope-labeled (SIL) amino acids and purified. Defined amounts of the SIL protein standard are added to the proteolytic digest of the sample, and endogenous and standard peptides (SIL) are released. Quantification is based on the ratios of labeled (standard) and unlabeled (sample) peptide(s). Re-quantified QconCAT protein standard enables absolute peptide quantification. Abbrev.: QconCAT: quantification concatemer; RVT: reverse translation; SIL: stable isotope-labeled; SILAC: stable isotope labeling by amino acids in cell culture; std: standard. Figure was created with BioRender. Figure is based and adapted from [206].

To attenuate this phenomenon, the use of short amino acid residue flanking sequences, mimicking the natural amino acid environment was proposed [221]. Unfortunately, this concept is only applicable to a limited extend, as its implementation results in a multiplication of QconCAT constructs that need to be designed, expressed, and quantified. An alternative approach enabling the high throughput generation of large amounts of peptide standards are Protein Epitope Signature Tags (PrESTs) [212]. PrESTs are short protein fragments representing unique regions of human proteins, discovered through the Human Protein Atlas project [211]. Utilizing this resource, stable isotope-labeled short protein sections (non-full length proteins) for absolute quantification can be generated in bulk amounts. A common issue associated with the use of individual peptide standards (such as AQUA peptides), non-full length protein standards (PrESTs) or artificial protein constructs (QconCATs) is their altered physicochemical behavior, which makes them incompatible with protein fractionation. Additionally, a noteworthy concern with using these strategies is that the protein quantification relies on only a few selected (typically one or two) peptides, which might compromise quantitative accuracy. Protein-based internal reference standards generated with the absolute

SILAC [213], FLEXIQuant [214] or PSAQ strategy [215] undeniably present an advantageous and appealing alternative. Utilizing the principle of metabolic labeling, full length proteins to be studied are expressed in an either cell-free (FLEXIQuant, absolute SILAC) or bacterial expression system (PSAQ, absolute SILAC). The purified and quantified protein standards can be added to the cell lysate, and can be further processed with the biological sample now allowing even for sample fractionation. Despite the benefits which full-length protein standards offer, the practicability of managing bulk expression, purification, and quantification is challenging and needs to be considered, especially when dealing with a larger number of target proteins.

2. Dissertation Rationale and Outline

The objective of this thesis was to develop and apply a mass spectrometry-based analysis platform that enables the unbiased and systematic determination of copy numbers for lysosomal proteins from different sample types of mouse origin.

Protein copy number information, which can only be derived from absolute quantitative data are scarce for lysosomal proteins and allow for the study and direct comparison of lysosomal protein amounts between cell/tissue types and conditions. Further lysosomal protein-, protein complex- and whole proteome stoichiometries can be determined. The implementation of an MS data acquisition strategy that is highly reproducible across various cell and tissue types of varying sample complexity, and does not rely on lysosome enrichment/isolation prior to MS expands the accessibility of samples to MS analysis. The presented work addresses the outlined demands, contributing to the broader aim of (absolutely) quantifying the dynamics in lysosomal proteome composition and determining individual protein expression levels across a range of cell/tissue types and experimental conditions.

In **Chapter 1**, the in-solution digestion using the MS-compatible surfactant RapiGest was optimized. The optimized protocol ensures the robust and universal applicability of RapiGest digestion for the samples of this project. **Chapter 2** focuses on establishing an MS data acquisition strategy, suitable for reproducible and accurate quantification of lysosomal proteins and lysosomal protein regulation in a complex background, such as whole cell and tissue lysate, without any prior enrichment. In this context, a modern hybrid MS approach (data-independent acquisition, DIA) combining features of untargeted and targeted MS, as well as a targeted MS approach (parallel reaction monitoring, PRM) were compared, exploring quantitative capabilities of both approaches in samples of varying complexity. In order to allow for the absolute quantification of lysosomal proteins in **Chapter 3**, a stable isotope-labeled, absolutely quantified internal standard using the so-called quantification concatemer (QconCAT) strategy was generated and applied. For targeted data acquisition, a multiple reaction monitoring (MRM) assay was developed. Absolute protein amounts enabled the determination of copy numbers for 143 proteins in one immortalized cell line, four primary cell types, and lysosome-enriched fractions.

3. Chapter 1 - Optimization of RapiGest In-Solution Digestion for Robust Sample Preparation

Based on: Mosen, P.R., R. Hardt, and D. Winter, *RapiGest precipitation depends on peptide concentration*. Proteomics, 2021. **21**(20): p. e2100129.
Online access via: https://doi.org/10.1002/pmic.202100129

3.1 Introduction

Bottom-up proteomics applies the enzymatic digestion of proteins into peptides which are separated through liquid chromatography and subsequently analyzed by mass spectrometry (MS) [78]. The accuracy and precision of the protein quantification rely on an efficient and reproducible release of peptides through protein digestion. In particular for absolute quantification, where the (endogenous) analyte signal is compared to a standard signal of known amount, complete proteolysis of the sample analyte and if applicable of the protein standard, is essential to achieve reliable quantifications [120, 220]. Therefore, to evaluate digestion efficiencies, besides the number of identified peptide and proteins [120, 222, 223] also the rate of missed cleaved peptides, respectively the ratio to the non-missed cleaved peptides should be considered in a quantitative study [120, 220, 222]. Numerous digestion protocols as well as their optimized variations exist and have been evaluated in this respect. Broadly, the available protocols can be categorized into three categories: in gel digestion, IGD [224]; on surface digestion either on beads (single-pot solid-phase-enhanced sample preparation, SP3 [225, 226]) or membranes [227]; in solution digestion either in tubes [222, 223], spin filters (filter aided sample preparation, FASP [228] or suspension trapping, STrap [229]) or tips (in-StageTip, iST [230]). In-solution digestion using the MS-compatible surfactant RapiGest [231, 232] consistently yields high(est) numbers of identified unique proteins and peptides, demonstrated excellent solubilization properties and provides low rates of missed cleavage [222, 223]. Due to these beneficial features, RapiGest was chosen for sample preparation/generation in this study. However, RapiGest's structural similarity to sodium dodecyl sulfate (SDS) requires acid hydrolysis of the intact RapiGest molecule and the removal of its hydrophobic moiety through precipitation after protein digestion [232]. Considerable variability was observed at the step of RapiGest precipitation for samples of different origin, resulting in an incomplete removal of the surfactant's hydrophobic moiety. Since this interference prevents a uniform and streamlined sample preparation and might also impede subsequent data acquisition, the underlying conditions behind the precipitation inconsistencies were examined, and a strategy was devised for the consistent removal of RapiGest.

3.2 Publication: Main Article

Received: 21 May 2021 | Revised: 5 August 2021 | Accepted: 25 August 2021 | Accepted article online: 27 August 2021

DOI: 10.1002/pmic.202100129

Proteomics
Proteomics and Systems Biology

TECHNICAL BRIEF

RapiGest precipitation depends on peptide concentration

Peter R. Mosen | **Robert Hardt** | **Dominic Winter**

Institute for Biochemistry and Molecular Biology, Medical Faculty, University of Bonn, Bonn, North Rhine-Westphalia, Germany

Correspondence
Dominic Winter, Institute for Biochemistry and Molecular Biology, Medical Faculty, Rheinische Friedrich-Wilhelms-University of Bonn, Nussallee 11, 53115 Bonn, NRW, Germany.
Email: dominic.winter@uni-bonn.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: FOR2625

Abstract

The mass spectrometry-compatible surfactant RapiGest promotes the enzymatic digestion of proteins by facilitating their unfolding while retaining enzymatic activity. RapiGest consists of a hydrophilic head and a hydrophobic tail, which can be separated by acid hydrolysis. This allows for removal of RapiGest prior to mass spectrometric analysis via precipitation and solid phase extraction. During in-solution digestion experiments with RapiGest, we noticed a high variability in the formation of precipitates after acid hydrolysis, implying that RapiGest precipitation is sample-dependent. We show that RapiGest hydrolyses efficiently under acidic conditions and that differences in precipitation are solely due to protein/peptide concentration. Furthermore, we demonstrate that RapiGest precipitation can be triggered by the addition of intact proteins, providing a strategy for its efficient removal from highly diluted samples. Data are available via ProteomeXchange with identifier PXD025982.

KEYWORDS
in solution digestion, precipitation, proteomics, RapiGest, surfactant

1 | INTRODUCTION

For the analysis of proteins by mass spectrometry (MS)-based proteomics, the vast majority of studies utilizes the so-called bottom-up approach, for which one of the key steps is the generation of peptides by proteolytic digestion [1]. Over the years, a large variety of digestion strategies has been developed. In principle, three types of approaches can be distinguished: in gel digestion [2], on surface digestion (beads, e.g., single-pot solid-phase-enhanced sample preparation, SP3 [3]; or membranes [4]), and in solution digestion (either in tubes [5]; spin filters, e.g., filter aided sample preparation, FASP [6]; or tips, e.g., in-StageTip, iST [7]). Especially for the preparation of complex samples, such as whole cell lysates, on bead, and in solution digestion approaches are used frequently. For such experiments, many protocols include a step for denaturing of samples in order to provide proteases access to the respective cleavage sites. A crucial factor for these protocols is that the compounds used for denaturation must not interfere with the following MS analyses [3,6]. Therefore, in solution digestion strategies frequently employ denaturing agents which can be separated from the generated peptides. Commonly used agents are either chaotropes, such as urea [5,8–10] or guanidinium chloride (GCl) [9], as well as MS-compatible removable surfactants, such as deoxycholate [8,9], ProteaseMAX [5,11], or RapiGest [5,8,10,12]. The latter are structurally similar to sodium dodecyl sulphate (SDS), being composed of a hydrophobic tail and a polar head group [11,12]. The separation of samples and MS-compatible surfactants is based on distinct processing steps. While deoxycholate can be removed by liquid phase extraction [9,13], ProteaseMax and RapiGest require a step that separates their hydrophobic and hydrophilic moieties, which is achieved by tryptic cleavage [11] or acid hydrolysis [12], respectively. Subsequently, the hydrophobic tail of the corresponding surfactant

Abbreviations: ABC, ammonium bicarbonate; AcOH, acetic acid; BLG, β-lactoglobulin; CAH, carbonic anhydrase; CASA, α-casein; CASB, β-casein; FA, formic acid; FASP, filter aided sample preparation; GCl, guanidium hydrochloride; HCD, higher-energy C-trap dissociation; iST, in-StageTip; LC-MSMS, liquid chromatography tandem mass spectrometry; MS, mass spectrometry; RT, room temperature; SP3, single-pot solid-phase-enhanced sample preparation; UHPLC, ultra-high performance liquid chromatography

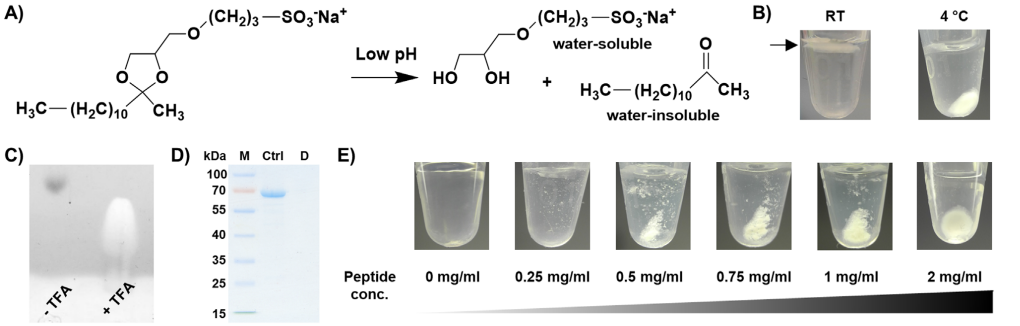


FIGURE 1 (A) Acid hydrolysis of RapiGest. (B) Formation of a floating disk (at RT, indicated by arrow) or a pellet (at 4°C) by centrifugation of tryptic BSA digests containing hydrolysed RapiGest. (C) Thin layer chromatography analysis of intact and hydrolyzed RapiGest. (D) SDS-PAGE analysis of digestion efficiency of a RapiGest BSA digest. (E) Precipitation of hydrolysed RapiGest in the presence of different concentrations of BSA peptides at 4°C. RT, room temperature; M, marker; Ctrl, non-digested; D, digested; BSA, bovine serum albumin; conc., concentration

can be precipitated by centrifugation and the hydrophilic headgroup is removed by reversed phase chromatography [11,12].

During the tryptic digestion of protein samples with RapiGest, we observed that acid hydrolysis and centrifugation does not always result in the formation of a pellet at the bottom of the sample tube. The precipitate, which consists of the hydrophobic moiety of the molecule (Figure 1A) [12], was sometimes not visible, while, in other cases, it was either floating on top of the sample or distributed along the tube wall.

As we always used the same concentration of RapiGest, we investigated several parameters that could influence RapiGest precipitation. Concerning the location of precipitated RapiGest on the top, middle, or bottom of the tube, we were able to identify the sample temperature as the decisive factor. Acid hydrolysis at 1% TFA for 45 min and 37°C, followed by centrifugation at 20,000 × *g* for 10 min at room temperature (RT), often resulted in a floating layer on top of the sample or in dispersed small pieces. Incubation of those samples for 45 min on ice and centrifugation at 20,000 × *g* for 10 min at 4°C, on the other hand, resulted in a solid pellet at the bottom of the tube (Figure 1B). For the complete lack of precipitation, we argued that either incomplete RapiGest hydrolysis or precipitation could be the reason. First, we tested whether quantitative hydrolysis of RapiGest was achieved. Therefore, we selected the most extreme situation, a solution of 0.1% RapiGest, 100 mM NH_4HCO_3 pH 7.8 (ABC) without any peptides and performed acid hydrolysis as well as centrifugation. This did not result in the formation of any pellet (or oily layer on top of the sample). To assess RapiGest cleavage in this sample, we analysed it by thin layer chromatography (TLC) on a heat-activated silica plate (silica gel 60 HPTLC, Merck, Darmstadt, Germany) at RT, using chloroform/methanol/water (70:30:4, v/v/v), followed by iodione staining. This analysis revealed complete cleavage of RapiGest, implying that only its precipitation is decisive for the observed effect (Figure 1C).

We subsequently evaluated several parameters, which may affect the formation of a RapiGest pellet using a proteolytic bovine serum albumin (BSA, Carl Roth, Karlsruhe, Germany) digest. For this, we dissolved BSA in 0.13% RapiGest (Waters, Milford, MA, USA), 100 mM ABC at a concentration of 2.6 mg/mL, reduced and alkylated disulphide bonds with 5 mM DTT (45 min at 56°C) and 20 mM acrylamide (30 min at RT), respectively, and quenched the reaction with 5 mM DTT (15 min at RT). After addition of trypsin (Promega, Mannheim, Germany) at an enzyme to sample ratio of 1:50 (w/w), samples were digested at a final protein concentration of 2 mg/mL and RapiGest concentration of 0.1% at 37°C overnight. Complete digestion was confirmed by combination of the sample with Laemmli buffer (1x final concentration) [14], SDS-PAGE, and staining of gels with Coomassie Brilliant Blue (Page Blue Protein Staining Solution, Thermo Fisher, Waltham, MA, USA) (Figure 1D). Using this sample, we tested different buffer (50, 100, and 200 mM ABC) and acid concentrations (0.5%, 1%, and 2% TFA), as well as differences in sample viscosity/density (0%, 5%, 10%, and 20% ACN). For all conditions, we did not observe any effects on RapiGest precipitation. Finally, we tested the effect of varying peptide concentrations obtained by dilution of the RapiGest digest with a 0.1% RapiGest solution (final concentration of 0.1 to 2 mg/mL), which showed a clear correlation with RapiGest pellet sizes (Figure 1E).

This effect may be related to an interaction of peptides and the hydrophobic tail of RapiGest, implying the possibility of co-precipitation, which was suggested previously [12]. In such a scenario, increasing amounts of RapiGest should co-precipitate increasing amounts of peptides, resulting in a reduction of peptide amount in the clear supernatant. To investigate potential peptide losses, we precipitated increasing amounts of RapiGest in a constant amount of an equimolar mixture of the trypsin-digested standard proteins BSA, β-lactoglobulin (BLG), ovalbumin (OVA), carbonic anhydrase (CAH), α-casein (CASA), and β-casein (CASB, all from Sigma-Aldrich, St. Louis, MO, USA). To exclude any effects due to the RapiGest concentration during proteolytic digestion, we first generated peptides separately by an alternative method, digestion in 0.8 M urea, 100 mM HEPES pH 7.5 (1 mg/mL protein concentration) at an enzyme to sample ratio of 1:50 (w/w). After desalting of peptides (3cc SepPak C_{18} cartridges, Waters), dried eluate fractions were resuspended in different concentrations of RapiGest (0.1%, 0.5%, 1%, and 2%), 100 mM ABC pH 7.8. Following

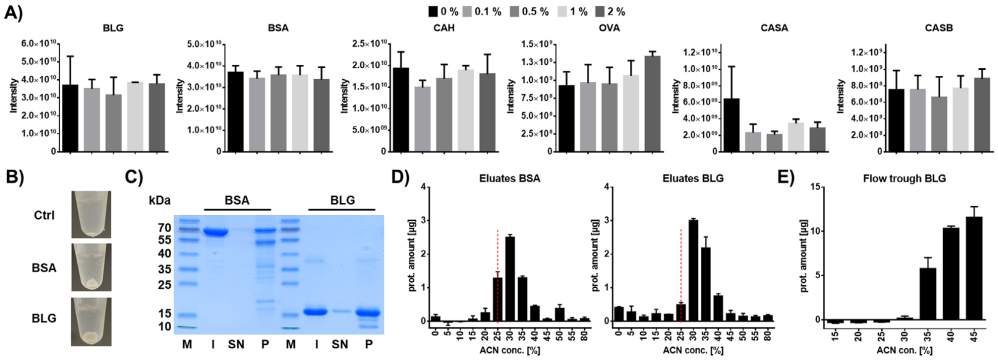


FIGURE 2 (A) Summed peptide intensities for the individual proteins of a six standard protein digest after precipitation with varying RapiGest concentrations determined by LC-MSMS. Shown are mean values ($n = 3$) + SD. (B) Induction of RapiGest precipitation by addition of BSA or BLG as trigger protein. (C) SDS-PAGE of individual fractions from RapiGest precipitation samples including BSA or BLG. (D) Elution of BSA and BLG from C_8 material with increasing ACN concentrations. Shown are mean values ($n = 2$) + SD. (E) Binding behaviour of BLG to C_8 stationary phase in varying amounts of ACN. Shown are mean values ($n = 2$) + SD. BLG, β-lactoglobulin; BSA, bovine serum albumin; CAH, carbonic anhydrase; OVA, ovalbumin; CASA, α-casein; CASB, β-casein; Ctrl, Control; M, marker; I, input; SN, supernatant; P, pellet; ACN, acetonitrile

RapiGest precipitation with 1% TFA and desalting of the clear supernatants using StageTips [15], we analysed the samples by LC-MSMS.

Measurements were performed with a Dionex Ultimate 3000 nano UHPLC system coupled to an Orbitrap Fusion Lumos mass spectrometer (both Thermo Fisher Scientific, Bremen, Germany) using in house-prepared 40 cm × 100 μm C_{18} analytical columns [10]. Peptides were loaded at a flow rate of 600 nL/min with 100% solvent A (water with 0.1% FA) and eluted with a 30 min linear gradient from 1% solvent B (90% ACN, 0.1% FA) to 35% solvent B at a flow rate of 300 nL/min. Eluting peptides were ionized in the positive ion mode and all scans performed in the Orbitrap analyser. After each survey scan (330–1500 m/z, resolution of 60,000), the most abundant peptides (m/z 2–7) were isolated at m/z 1.4 in the quadrupole, fragmented by HCD (30% collision energy), and analysed in the top speed mode with a cycle time of 2 s (fragment ion spectrum resolution of 30,000) and dynamic exclusion of 15 s. Acquired data were analysed with MaxQuant 1.6.5 [16], utilizing a database containing the six standard proteins in combination with the default contaminants list. Oxidation (M), acetylation (protein N-terminus), and deamidation (NQ) were set as variable modifications, and propionamide (C) was defined as fixed modification. Trypsin/P was set as enzyme, up to two missed cleavages were allowed, and standard settings were used for mass tolerances. Due to the small size of the database, FDR filtering was deactivated and a fixed Andromeda score cut-off of 40 was applied.

Comparison of the summed peptide intensities of the individual proteins did not reveal strong effects between the different conditions, with the exception of CASA, for which we observed for all RapiGest concentrations a trend towards decreased signal intensities relative to the control sample (Figure 2A, Table S1). This effect could be related to a stronger interaction of CASA-derived peptides with the hydrophobic part of RapiGest, resulting in their co-precipitation.

The lack of RapiGest precipitation in low concentrated samples suggests that sufficient reduction of sample volume, and therefore an increase of peptide concentration to at least 0.5 mg/mL (Figure 1E), could allow for efficient precipitation. As volume reduction can only be performed to a certain extend before it becomes impracticable to separate the RapiGest pellet from the supernatant, this strategy is, however, limited by the sample amount. Alternatively, peptide concentration could be increased by the addition of external peptides. Such peptides would, however, most likely mask low abundant co-eluting sample peptides, and block a significant amount of column binding capacity. This is, however not the case for intact proteins, as they can be separated from peptides based on their chemical properties.

We therefore tested the removal of hydrolysed RapiGest by adding intact "trigger proteins". In these experiments, we added either BSA or BLG to a low concentrated tryptic BSA digest (peptide concentration of 0.25 mg/mL), which previously did not result in the formation of a proper pellet (Figure 2B, top), followed by acid hydrolysis of RapiGest and centrifugation. Both supplementation with BSA and BLG (1 mg/mL final concentration), which differ in size and hydrophobicity (GRAVY scores [17]/molecular weights of -0.43/69 kDa and -0.01/20 kDa, respectively), resulted in the precipitation of RapiGest with a slightly bigger pellet in the case of BLG (Figure 2B). Investigation of the protein distribution in the individual fractions of these samples by SDS-PAGE showed that the majority of intact proteins was present in the pellet (Figure 2C), implying that most of the trigger protein is removed during RapiGest precipitation [12]. We did, however, observe residual leftovers of trigger protein in the supernatant (Figure 2C, middle lanes).

As the presence of residual intact trigger proteins in a low concentrated peptide sample could interfere with its subsequent LC-MSMS analysis, we directed our attention towards its removal. In theory,

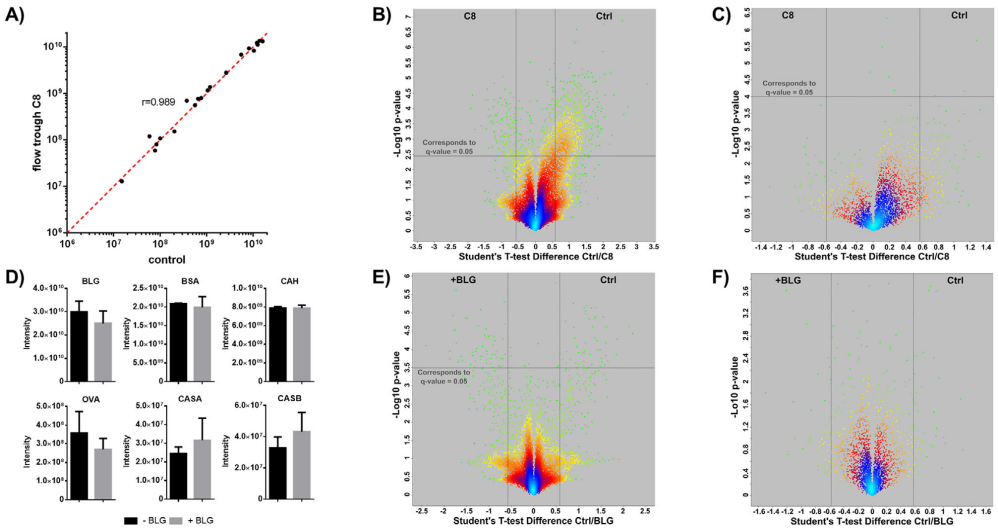


FIGURE 3 (A) Correlation of BSA peptide intensities derived from LC-MSMS analyses of a control and C_8 flowthrough sample. (B, C) Volcano plot of Student's *t*-test result ($n = 3$, Benjamini-Hochberg FDR = 0.05). Peptide (B) or protein (C) abundance differences between control and C_8 flowthrough HeLa samples are plotted against − log10 *p*-values. Points are coloured according to their Gaussian kernel density. (D) Summed peptide abundances for the individual proteins of a six-protein digest in which RapiGest was precipitated with/without BLG. Shown are mean values of $n = 3$ + SD. (E, F) Volcano plot of Student's *t*-test result ($n = 3$, Benjamini-Hochberg FDR = 0.05). Peptide (E) or protein (F) abundance differences between control and BLG supplemented HeLa samples are plotted against − log10 *p*-values. Points are coloured according to their Gaussian kernel density. Ctrl, control; BSA, bovine serum albumin; BLG, β-lactoglobulin; OVA, ovalbumin; CAH, carbonic anhydrase; CASA, α-casein; CASB, β-casein

proteins should still be retained by C_8 reversed phase material at ACN concentrations where the vast majority of peptides is found in the flowthrough. We therefore determined the maximum ACN concentration at which BSA and BLG are retained by C_8 tip columns. After binding of proteins resuspended in 100 mM ABC, 1% TFA to Empore C_8 StageTips (3M, Saint Paul, MN, USA), we eluted the tips with increasing concentrations of ACN/0.5% acetic acid (AcOH). Eluate fractions were dried using a vacuum centrifuge, resuspended in 4% SDS, and the protein content was determined using the DC protein assay (Bio-Rad, Hercules, CA, USA). For BSA, we observed an earlier elution (25% ACN) compared to BLG (30% ACN), indicating a weaker interaction with the stationary phase (Figure 2D). As BLG also resulted in more efficient precipitation of RapiGest (Figure 2B), we selected it as trigger protein. As not the elution of BLG from C_8 material but its ability to bind to it at a certain concentration of ACN is decisive for its separation from peptide samples, we further tested at which maximal ACN level BLG is still retained by C_8 material. We loaded BLG, which was dissolved in buffers containing different amounts of ACN, on C_8 StageTips and quantified protein levels in the flowthrough fractions. These experiments showed, that BLG is efficiently retained by C_8 material at sample ACN concentrations of up to 30% (Figure 2E).

We subsequently assessed possible peptide losses due to the C_8 material in samples of low and high complexity (tryptic digests of BSA and HeLa whole cell lysate). For this purpose, HeLa cells were cultured in DMEM supplemented with 10% FCS, penicillin (100 IU/mL)/streptomycin (100 μg/mL) and 2 mM L-glutamine, washed with 1x PBS, harvested by scraping, and centrifuged at 1000 × *g* for 5 min at 4°C. The supernatant was discarded, the pellet resuspended in 4% SDS, 100 mM HEPES pH 7.5, and cells were lyzed by incubation at 95°C for 10 min, sonication for 3 × 30 s at 60% duty cycle using an Ultrasonics Sonifier (Branson, Brookfield, CT, USA), and a second heating step. After centrifugation at 20,000 × *g* for 30 min at RT, the clear supernatant was recovered, the protein content determined, and a sample containing 1 mg of protein was transferred to a new microtube. Proteins were precipitated by chloroform-methanol (2:1 v/v) precipitation, and the air-dried protein pellet was solubilized in 0.3% RapiGest, 100 mM ABC pH 7.8 for 15 min at 95°C. Trypsin was added at an enzyme to sample ratio of 1:500 (w/w) and the sample was incubated for 60 min. Further sample processing was performed as previously described for the standard proteins and the digest was carried out overnight at a final concentration of 0.1% RapiGest.

After precipitation of RapiGest, supernatant volumes corresponding to either 100 fmol of BSA or 10 μg of HeLa digest were adjusted to 30% ACN and passed over a C_8 StageTip (50 fmol of BSA peptides and 2.5 μg of HeLa digest per C_8 layer, respectively). The flowthrough fraction was collected, desalted using C_{18} StageTips, and analysed by

LC-MSMS using a 30 or 120 min gradient for BSA and HeLa whole cell lysate digests, respectively. The BSA experiment was carried out once, HeLa experiments were performed in in three independent replicates. For data analysis of HeLa samples, Swiss-Prot *human* (release date: 2021-07 with 20,371 entries) was used for database searching, and results were filtered at an FDR of 0.01. For BSA, we plotted values for individual peptides between C_8 exposed and control samples, revealing an excellent correlation (Figure 3A). For HeLa samples, data were further processed with Perseus v1.6.15 [18]. We considered only peptides quantified with three valid values in at least one of the two conditions and performed missing value imputation based on a down-shifted normal distribution. We then performed a two-sided Student's *t*-test with an FDR of 0.05 and applied a fold-change cut-off of 1.5-fold for the determination of up-/downregulated peptides (Table S1). On the peptide level, we observed losses in signal intensity for ~ 4% of peptides in the sample exposed to C_8 material while ~ 1% of peptides showed higher values (Figure 3B). This translated on the protein level to significant changes for 3 out of 3497 proteins (Figure 3C). Furthermore, we observed that the dataset has an asymmetric distribution for a subpopulation of peptides, whose change in signal intensity is below the applied cut-off. These data imply, that the exposure of the sample to C_8 material under the experimental conditions used (2.5 μg of peptides/disk in 30% ACN) do result in certain peptide losses. Therefore, it should be carefully considered if a C_8 material clean-up is necessary, for example, by analysis of a small fraction of the sample by SDS-PAGE and silver staining, as residual amounts of trigger protein will most likely not have a negative effect on the subsequent LC-MSMS analyses.

We performed the same experiment for the addition of trigger proteins, to address if the addition of BLG also results in peptide losses due to an increased co-precipitation of peptides with RapiGest. Utilizing the urea digested mixture of six standard proteins as well as the HeLa whole cell lysate digest, we precipitated RapiGest with/without addition of BLG, followed by C_8 StageTip treatment and LC-MSMS analysis. No statistically significant peptide loss was observed for the six standard protein digest (Figure 3D, Table S1). We did, however, observe trends opposing the results obtained for precipitation of the six standard proteins digest without addition of BLG as trigger protein (Figure 2A), with a slight decrease in signal intensity for BLG and OVA, and an increase for CASA and CASB. For CASA, the increase in signal intensity might result from hydrolysed RapiGest, which does not co-precipitate with CASA peptides anymore but rather with intact BLG, resulting in a better recovery of these peptides. For the HeLa whole cell lysate digest, we did not observe a trend towards loss or gain of signal intensity for individual peptides (54 and 57 out of 18,401 peptides had a lower/higher signal intensity due to the addition of BLG) and no changes on the protein level (Figure 3E,F). Furthermore, we did not observe any trends in sample distribution, as it was the case for exposure to the C_8 material, indicating that the addition of BLG as trigger protein has no negative effects on the sample.

In conclusion, we show that the precipitation of the acid-cleavable surfactant RapiGest depends on the peptide/protein concentration of the sample. To ensure proper pellet formation, we advocate for a peptide concentration of at least 1 mg/mL. For lower concentrated samples, we propose the addition of trigger proteins to the digest, which co-precipitate with the hydrophobic RapiGest cleavage product. This provides a strategy to reliably induce RapiGest precipitation, and to remove it from low concentrated samples.

ACKNOWLEDGMENTS

The authors thank Norbert Roesel for technical support. The workflow presented in the Table of Contents graphic was created with BioRender.com and the chemical structure was created with ChemDraw 16, a registered trademark of PerkinElmer Informatics. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD025982. [19].

Open access funding enabled and organized by Projekt DEAL.

AUTHOR CONTRIBUTIONS

Dominic Winter conceived the study and designed the experiments. Peter R. Mosen and Robert Hardt performed the experiments, acquired and analysed the data. Dominic Winter drafted the manuscript. Dominic Winter, Peter R. Mosen and Robert Hardt revised the manuscript. All authors have given approval to the final version of the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in ProteomeXchange Consortium via the PRIDE partner repository at https://www.ebi.ac.uk/pride/, reference number PXD025982.

ORCID

Peter R. Mosen https://orcid.org/0000-0001-5922-3805
Robert Hardt https://orcid.org/0000-0003-2939-7657
Dominic Winter https://orcid.org/0000-0001-6788-6641

REFERENCES

1. Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature, 537* (7620), 347–355.
2. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., & Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocol, 1* (6), 2856–2860.
3. Hughes, C. S., Foehr, S., Garfield, D. A., Furlong, E. E., Steinmetz, L. M., & Krijgsveld, J. (2014). Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular Systems Biology*, *10* (10), 14.
4. Bennike, T. B., Bellin, M. D., Xuan, Y., Stensballe, A., Møller, F. T., Beilman, G. J., Levy, O., Cruz-Monserrate, Z., Andersen, V., Steen, J., Conwell, D. L., & Steen, H. (2018). A cost-effective high-throughput plasma and serum proteomics workflow enables mapping of the molecular impact of total pancreatectomy with islet autotransplantation. *Journal of Proteome Research*, *17* (5), 1983–1992.
5. Winter, D., & Steen, H. (2011). Optimization of cell lysis and protein digestion protocols for the analysis of HeLa S3 cells by LC-MS/MS. *Proteomics*, *11* (24), 4726–30.

6. Wisniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods*, *6* (5), 359–62.
7. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., & Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature Methods*, *11* (3), 319–324.
8. Leon, I. R., Schwammle, V., Jensen, O. N., & Sprenger, R. R. (2013). Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis. *Molecular & Cellular Proteomics*, *12* (10), 2992–3005.
9. Proc, J. L., Kuzyk, M. A., Hardie, D. B., Yang, J., Smith, D. S., Jackson, A. M., Parker, C. E., & Borchers, C. H. (2010). A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. *Journal of Proteome Research*, *9* (10), 5422–5437.
10. Ponnaiyan, S., Akter, F., Singh, J., & Winter, D. (2020). Comprehensive draft of the mouse embryonic fibroblast lysosomal proteome by mass spectrometry based proteomics. *Scientific Data*, *7* (1), 68–81.
11. Saveliev, S. V., Woodroofe, C. C., Sabat, G., Adams, C. M., Klaubert, D., Wood, K., & Urh, M. (2013). Mass spectrometry compatible surfactant for optimized in-gel protein digestion. *Analytical Chemistry*, *85* (2), 907–914.
12. Yu, Y. Q., Gilar, M., Lee, P. J., Bouvier, E. S. P., & Gebler, J. C. (2003). Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Analytical Chemistry*, *75* (21), 6023–6028.
13. Masuda, X. T., Tomita, T., & Ishihama, Y. (2008). Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *Journal of Proteome Research*, *7* (2), 731–740.
14. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage-T4. *Nature*, *227* (5259), 680–685.
15. Rappsilber, J., Ishihama, Y., & Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical Chemistry*, *75* (3), 663–670.
16. Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26* (12), 1367–72.
17. Kyte, J., & Doolittle, R. F. (1982). A Simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157* (1), 105–132.
18. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., & Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, *13* (9), 731–740.
19. Vizcaíno, J. A., Deutsch, E., Wang, R., Csordas, A., Reisinger, F., Ríos, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H., Albar, J. P., & Hermjakob, H. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnology*, *32* (3), 223–226.

SUPPORTING INFORMATION

Additional supporting information may be found online https://doi.org/10.1002/pmic.202100129 in the Supporting Information section at the end of the article.

How to cite this article: Mosen, P. R., Hardt, R., & Winter, D. (2021). RapiGest Precipitation Depends on Peptide Concentration. *Proteomics*, *21*, e2100129. https://doi.org/10.1002/pmic.202100129

3.3 Conclusion

To overcome the discovered, seemingly erratic precipitation behaviour of RapiGest [232], the factors contributing to the surfactant's successful precipitation and removal were investigated. Among the tested sample parameters which encompass among others buffer and acid concentration as well as sample viscosity, the main factor influencing RapiGest precipitation is the peptide concentration. Increasing peptide concentration positively correlated with the size of the RapiGest pellet (Figure 1E). As the results suggested a co-precipitation of peptides with the hydrophobic tail of RapiGest [232], a potential peptide loss was assessed. Therefore, a digest of an equimolar mix of six standard proteins was precipitated with increasing amounts of RapiGest, since this should result in an increasing loss of peptides (Figure 2A). No systematic trend for the (co-) depletion of peptides was observed, indicating that there is no substantial peptide loss to be anticipated when using RapiGest. Based on experimental experience and application of RapiGest, a minimum peptide/protein concentration of 1 µg/µL is recommended to ensure proper pellet formation. For samples with low(er) peptide concentration, like those, e.g., derived from a FACS (Fluorescence-Activated Cell Sorting) experiment, reducing sample volume is only feasible up to a certain extent before preparation of samples using RapiGest becomes impracticable. In such a scenario, RapiGest precipitation can be reproducibly achieved by addition of a trigger protein such a β-lactoglobulin, leading to the co-precipitation of the majority of intact trigger protein with RapiGest (Figure 2B-C). For a further, optional removal of remaining intact trigger protein, a strategy involving C_8 reversed-phase material was devised. Here, the acidified sample digest with the spiked trigger protein (β-lactoglobulin) is adjusted to 30% acetonitrile (ACN) and passed over the C_8-StageTips. While residual trigger protein is retained, peptides can pass and can be collected from the flowthrough for subsequent MS sample preparation (Figure 2E). The impact of C_8 material exposure, along with the co-precipitation of β-lactoglobulin on peptides, was examined using HeLa digests. Although the addition and co-precipitation of the trigger protein with RapiGest itself did not result in sample losses, the subsequent exposure of the sample to C_8 material for removal of the residual trigger protein might lead to certain peptide losses, which however only had a minimal impact on final protein quantification (Figure 3E-F).

4. Chapter 2 - Comparative Assessment of DIA and PRM Mass Spectrometry for Accurate Quantitative Analysis of the Lysosomal Proteome in Complex Samples

Based on: Mosen, P., et al., *Targeted Quantification of the Lysosomal Proteome in Complex Samples*. Proteomes, 2021. **9**(1): p. 9010004.

Remark: The presented research was conducted in collaboration with my co-first author, Anne Sanner, reflecting an equally shared contribution. For this work, I developed a targeted MS assay (PRM) and devised the parameters as well as the strategy for the streamlined analysis of large-scale targeted mass spectrometry data. I prepared biological samples, acquired the PRM-MS data, and curated the datasets generated during the course of this study. Furthermore, I analyzed and compared the two datasets (PRM and DIA) of this study and visualized the acquired data. Together with the other authors I wrote, edited, and reviewed the manuscript and worked on its revision.

Online access via: https://doi.org/10.3390/proteomes9010004

4.1 Introduction

The intended, standard-based absolute quantification of low abundant lysosomal proteins from whole cell and tissue lysates, without any enrichment prior to MS, presents numerous challenges to data acquisition. Classic, untargeted DDA suffers from a significantly reduced reproducibility of peptide identification and quantification [89]. Caused by the semirandom sampling of peptides in DDA mode, this results mainly in the coverage of the most abundant peptides per sample [93, 233], thereby deeming DDA-MS unsuitable for the project's scope. To overcome quantification issues associated to DDA, data acquisition strategies such as MRM and PRM specifically target a set of user-defined, pre-selected peptides of interest [126, 131]. Here, the acquisition of data occurs independently of peptide (ion) abundance, ensuring a highly reproducible and accurate quantification of even lowest amounts of analyte [120, 121]. However, its implementation demands manual assay design and development, involving the specification of peptide m/z and, in case of larger proteomics studies, also the definition of retention times for optimal assay performance [132]. The DIA approach [90, 234] can be considered, depending on the particular experimental setup as a hybrid data acquisition method [235]. Combining features from both, untargeted data acquisition involving the continuous and unbiased fragmentation of all peptides precursor ions within a specific m/z range, as well as targeted MS data extraction for data analysis [234, 236], data-independent acquisition should provide a level of reproducibility comparable to targeted proteomics, yet its data acquisition is not constrained to a defined set of peptides and thus does not require any assay development [235]. While DIA has demonstrated superiority in several aspects when compared to DDA [54, 93, 236, 237], performance comparisons between DIA and targeted MS methods such as MRM [92] and PRM [238] are rare. Furthermore, such comparisons exclusively date back to the early stages of DIA

(around the year 2015) when DIA-MS - as we know it today [239] - began gaining wider recognition within the proteomics community. DIA has become, next to DDA a widely used data acquisition strategy which is applied by now to all types of samples with varying complexity [240], ranging from low complex isolate samples [56, 241, 242] and medium complex cell sample [56, 243] to high complex tissue samples [54, 244, 245]. In general, DIA is recognized and promoted as data acquisition strategy that provides (almost [246]) quantitative characteristics of targeted proteomics for features such as sensitivity, accuracy and selectivity at a large scale [247], thereby presumably enabling deep proteome coverage [240, 248]. However, already the initial comparative DIA studies mainly performed in medium complex lysates of yeast [90, 238], cell lines [90, 238, 249] as well as in human plasma [250] indicated a performance difference in sensitivity (between 3- to 10-fold) compared to traditional targeted MRM-MS [91]. It should be noted that the initial implementation of the DIA-MS approach (sequential window acquisition of all theoretical mass spectra, SWATH-MS [90]) was carried out on fast-scanning tripleTOF-instruments [73], which offer a lower sensitivity compared to triple quadrupole instruments [251, 252]. Since this beginning DIA-MS has evolved, was implemented on Orbitrap (OT)-instruments [253] and is now predominantly performed on these [248, 254, 255]. Recent comparative studies between DIA-MS conducted on OT-instruments and MRM- or PRM-MS are rare. Kockmann et al. [238] concluded that targeted data acquisition strategies (MRM and PRM) consistently demonstrate superior quantitative accuracy and precision compared to data-independent acquisition (DIA) methods. This advantage became even more evident when the analyte concentration was reduced to low levels. As previously mentioned, performance comparisons for quantification using DIA were conducted mostly in yeast and cell lysates (samples of medium complexity).

For the study of LSDs, however, animal-derived models and consequently tissue-based samples are of primary interest. Furthermore, the reviewed and aforementioned studies did not include lysosomal proteins/peptides in their quantitative performance evaluation (between DIA and MRM/PRM), and peptide selection seemed rather random and sporadic. As a result, based on the reviewed literature no conclusive answer regarding the suitability of DIA for targeted detection of low abundant lysosomal proteins was possible. For the standard-based absolute quantification of low-abundant lysosomal proteins, achieving reproducible and accurate quantification is essential. Data-independent acquisition (DIA), with its unbiased and algorithm-controlled acquisition mode eliminates the need for manual assay development and thereby presents an intriguing alternative to traditional targeted MS data acquisition strategies that rely on sophisticated assay design and optimization. A comparative analysis of quantitative performance between data-independent acquisition (DIA)-MS and parallel reaction monitoring (PRM) MS on an Orbitrap Fusion Lumos system was conducted, assessing the quantitative capabilities of each method in analyzing the lysosomal proteome across samples of varying complexities.

4.2 Publication: Main Article

proteomes

MDPI

Article

Targeted Quantification of the Lysosomal Proteome in Complex Samples

Peter Mosen †, Anne Sanner †, Jasjot Singh and Dominic Winter *

Institute for Biochemistry and Molecular Biology, Medical Faculty, University of Bonn, 53115 Bonn, Germany; pmos@uni-bonn.de (P.M.); anne.sanner@uni-bonn.de (A.S.); jsin@uni-bonn.de (J.S.)
* Correspondence: dominic.winter@uni-bonn.de; Tel.: +49-228-73-7081
† These authors contributed equally.

Abstract: In eukaryotic cells, lysosomes play a crucial role in the breakdown of a variety of components ranging from small molecules to complex structures, ascertaining the continuous turnover of cellular building blocks. Furthermore, they act as a regulatory hub for metabolism, being crucially involved in the regulation of major signaling pathways. Currently, ~450 lysosomal proteins can be reproducibly identified in a single cell line by mass spectrometry, most of which are low-abundant, restricting their unbiased proteomic analysis to lysosome-enriched fractions. In the current study, we applied two strategies for the targeted investigation of the lysosomal proteome in complex samples: data-independent acquisition (DIA) and parallel reaction monitoring (PRM). Using a lysosome-enriched fraction, mouse embryonic fibroblast whole cell lysate, and mouse liver whole tissue lysate, we investigated the capabilities of DIA and PRM to investigate the lysosomal proteome. While both approaches identified and quantified lysosomal proteins in all sample types, and their data largely correlated, DIA identified on average more proteins, especially for lower complex samples and longer chromatographic gradients. For the highly complex tissue sample and shorter gradients, however, PRM delivered a better performance regarding both identification and quantification of lysosomal proteins. All data are available via ProteomeXchange with identifier PXDD023278.

Keywords: targeted proteomics; lysosomes; parallel reaction monitoring; data-independent acquisition; label-free quantification

Citation: Mosen, P.; Sanner, A.; Singh, J.; Winter, D. Targeted Quantification of the Lysosomal Proteome in Complex Samples. *Proteomes* **2021**, *9*, 4. https://doi.org/10.3390/proteomes9010004

Academic Editor: Jörg Reinders
Received: 23 December 2020
Accepted: 21 January 2021
Published: 26 January 2021

1. Introduction

Lysosomes are membrane-bound organelles, which are well-known as the main degradative compartment of eukaryotic cells [1]. They fulfil a crucial function for the breakdown of a variety of cellular components and the recycling of their building blocks. This is achieved by ~60 hydrolases and ~40 transporters residing in the lysosomal lumen and membrane [2]. The proper function of these hydrolases is crucial for cellular homeostasis, as exemplified by the detrimental consequences of lysosomal enzyme malfunction. Mutations resulting in their altered activity, stability, or subcellular distribution can result in the accumulation of their respective substrates within lysosomes, interfering with the correct function of the organelle. Impaired lysosomal function is the primary cause of a group of ~70 inherited rare genetic diseases, so-called lysosomal storage disorders (LSDs), which frequently result in neurodegeneration, metabolic dysfunction, impaired development, and premature death [3]. To date, therapies exist only for a handful of LSDs and those available are almost exclusively symptomatic [3–5].

While the connection between lysosomal dysfunction and LSDs has been known for decades, altered lysosomal or lysosome-associated proteins have recently been shown in an increasing number of studies to be involved in more common conditions, increasing the public interest in this organelle. This includes, but is not limited to, cancer [6], neurodegenerative disorders [7], and cardiovascular diseases [8]. As part of this development, the view on lysosomes as unregulated cellular waste bags, which persisted for decades, is currently

transitioning towards highly mobile organelles that act as a major regulatory hub of cellular metabolism. In recent years, lysosomes have been shown to vary in their properties, to be actively transported, to interact with other organelles, and to respond to various cellular and environmental stimuli with the help of an extensive network of proteins [2,9,10]. This involves several key players regulating cellular growth and energy metabolism, such as the mechanistic target of rapamycin complex 1 (mTORC1) or the AMP-dependent kinase (AMPK), which are activated at the lysosomal surface [11].

These emerging roles of lysosomes have led to an increasing interest in the analysis of lysosomal proteins. For the unbiased characterization of large numbers of proteins, mass spectrometry (MS)-based proteomics is currently the method of choice, as it allows for the identification, quantification, and characterization of thousands of proteins from a given sample [12]. To date, ~740 proteins have been assigned in one way or the other to lysosomes, ~300 of which are either located in the lysosomal lumen, at the lysosomal surface, or directly interact with it [13].

Lysosomal proteins are typically of low abundance and therefore frequently not covered in DDA whole proteome shotgun analyses. The most common way to increase the coverage of lysosomal proteins is lysosome enrichment, resulting in a reduced sample complexity and therefore facilitating their analysis. Several lysosome enrichment methods are currently available, the most common of which are based either on density gradient centrifugation, superparamagnetic iron oxide nanoparticles (SPIONs) in combination with magnetic columns, or the immunoaffinity enrichment of tagged lysosomal proteins [14]. While all of these approaches allow for a certain degree of enrichment, they come with restrictions that limit the samples they can be applied to. While density gradient centrifugation can be performed for virtually any starting material, recovery is low and high amounts of contaminating organelles, mostly mitochondria, are included in the lysosome-containing fractions [14]. The use of SPIONs, which are taken up by unspecific fluid phase endocytosis and delivered to the lysosomal compartment through the endocytic pathway [15], is limited to cells grown in culture which actively perform fluid phase endocytosis. Furthermore, with this approach, only those lysosomes can be isolated that receive cargo from late endosomes, which may be affected when studying models of LSDs with impaired endosome-lysosome fusion. For the immunoaffinity enrichment of lysosomes through tagged membrane proteins [16,17], a fusion protein has to be stably expressed in cells or animals, requiring the generation of the respective organism. Furthermore, only lysosomes expressing the protein of choice are covered, which may result in a selection bias, and the overexpression of the tagged protein may influence lysosomal properties. For all approaches, millions of cells or milligram amounts of tissue are needed as starting material, excluding low-abundant samples from these analyses. As many LSDs affect distinct populations of cells, and the majority of LSDs can only be simulated in animal models, these limitations stall the proteomic investigation of LSDs, as it is frequently not possible to obtain lysosome-enriched fractions in sufficient quantities.

The need for enrichment of lysosomal proteins arises from the limitations of untargeted data-dependent acquisition (DDA)-based acquisition strategies, as highly abundant peptides prevent the fragmentation, and therefore identification, of those originating from low-abundant lysosomal proteins. Therefore, a promising alternative for the characterization of lysosomal proteins from small amounts of complex samples are targeted proteomics strategies. Currently, two major approaches for targeted proteomics are applied. On the one hand, previously defined peptides are fragmented in single, multiple, or parallel reaction monitoring (SRM, MRM, PRM) experiments, and abundance is determined based on the intensity of their fragment ions [18]. On the other hand, unbiased fragmentation of pre-determined m/z windows is performed in data-independent acquisition (DIA) approaches, and the abundance of the respective peptides is determined from unique fragment ions identified in mixed MS/MS spectra [19]. In comparison to DDA-based label-free quantification strategies, PRM and DIA approaches offer increased sensitivity and reproducibility for low-abundant peptides in complex samples [20–22], making them

ideal candidates for the analysis of lysosomal proteins from cell or tissue samples without prior enrichment.

So far, to our knowledge, targeted approaches have not been frequently used for the investigation of the lysosomal proteome. PRM was applied in a few studies for the investigation of selected lysosomal proteins (e.g., [23–25]), while only DIA approaches have been used for the analysis of the whole lysosomal proteome, so far solely investigating lysosome-enriched samples [14,16,26]. While it was reported that DIA is able to identify and quantify > 10,000 proteins within a single run [27], the reproducible quantification of lysosomal proteins suffers in highly complex samples and the achievable performance in whole cell lysates is significantly lower compared to lysosome-enriched fractions [14].

In the present study, we compared DIA and PRM for the analysis of the lysosomal proteome from samples of different complexities. We investigated lysosome-enriched fractions, as well as whole cell and liver lysate, and systematically compared the performance of DIA and PRM. While we could detect lysosomal proteins with both approaches in all sample types, and DIA identified higher numbers for most samples, PRM showed a better performance in liver lysate allowing for the detection of quantitative changes which were not identified by DIA.

2. Material and Methods

2.1. Cell Culture Experiments and Sample Lysis

All cell culture experiments were performed under a sterile hood and all solutions were pre-warmed to 37 °C. Mouse embryonic fibroblasts (MEFs) were cultured at 37 °C and 5% CO_2 in Dulbecco's modified eagle medium (DMEM), supplemented with 10% (*v*/*v*) fetal calf serum (FCS), 100 IU/mL penicillin, and 100 µg/mL streptomycin. For the generation of MEF whole cell lysate samples, 1.5×10^6 cells each were seeded on three 15 cm plates and cultivated for 72 h. The cells were washed once with 5 mL of ice-cold 1× phosphate-buffered saline (PBS), scraped in 600 µL of ice-cold PBS, and collected in a 1.5 mL microtube. Cells were pelleted by centrifugation at 1000× *g* and 4 °C for 4 min, the supernatant was discarded, and the cell pellet was re-suspended in 600 µL lysis buffer (4% SDS, 100 mM HEPES pH 7.5). Subsequently, the cell suspension was incubated at 95 °C for 10 min followed by sonication using a Ultrasonics Sonifier 250 (Branson, Danbury, CT, USA) at a duty cycle of 60% and an output of 6 for 90 s. Samples were centrifuged at 20,000× *g* and RT for 30 min and the clear supernatants were transferred to new microtubes.

Lysosome isolation was performed from MEF cells using SPIONs as described elsewhere [26]. In brief, cells were cultivated in DMEM with 2.5% FCS for 72 h (3×10^6 cells per 10 cm dish), 1 mL of magnetite solution (EndoMAG40, Liquids Research, North Wales, UK) was added to each plate, and the cells were incubated for 24 h (pulse period). Subsequently, the cells were washed twice with pre-warmed PBS, fresh DMEM (10% FCS) was added, and the cells were incubated for 24 h (chase period). Cells were washed with ice-cold PBS and harvested using a cell scraper in 2 mL lysosome isolation buffer (250 mM sucrose, 10 mM HEPES pH 7.4, 1 mM $CaCl_2$, 15 mM KCl, 1 mM $MgCl_2$, 1.5 mM MgAc, 1 mM dithiothreitol (DTT), 1x cOmplete EDTA-free protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany)). Plasma membranes were disrupted using a dounce homogenizer, and lysosomes were enriched using Miltenyi LS columns (Miltenyi Biotech, Auburn, CA) and eluted using a plunger.

2.2. Preparation of Mouse Liver Samples

Mice were handled in accordance with local regulations concerning the welfare of animals. Three months-old male C57BL/6 mice were sacrificed by cervical dislocation, the liver was extracted, and snap-frozen in liquid nitrogen. The frozen tissue was chopped into small pieces using a razor blade, and 1 mL of lysis buffer (4% SDS, 100 mM HEPES pH 7.5) was added. The sample was incubated for 10 min at 95 °C and sonicated using an Ultrasonics Sonifier at a duty cycle of 60% and an output of 6 for 90 s. Subsequently,

the samples were again incubated at 95 °C for 10 min, centrifuged at 20,000× *g* and RT for 30 min, and the clear supernatants were transferred to new microtubes.

2.3. Sample Preparation for Mass Spectrometry

The protein concentration of all samples was determined using the DC Protein Assay (Bio-Rad Laboratories, CA, USA). For MEF whole cell lysate and liver samples, 100 µg of protein were used for each replicate while 20 µg were used for lysosome-enriched fractions. Sample volumes were adjusted to 200 µL using HPLC-grade water and proteins were precipitated by addition of 1 mL ice-cold chloroform/methanol (2:1 *v*/*v*), vigorous vortexing, and centrifugation at 20,000× *g*, 4 °C for 1 h. The liquid phases were discarded, the protein pellets washed once with 1 mL of ice-cold methanol, and centrifuged at 20,000 × *g*, 4 °C for 15 min, followed by the removal of methanol. Protein pellets were air-dried and solubilized in 1% RapiGest (Waters, Milford, MA, USA), 0.1 M NH_4HCO_3 pH 7.8 at 95 °C for 10 min. Subsequently, samples were diluted 1 to 5 with 0.1 M NH_4HCO_3 and trypsin (Promega, Mannheim, Germany) was added at an enzyme-to-substrate ratio of 1:500, followed by incubation at 37 °C, 800 rpm in a thermomixer for 45 min. Proteins were reduced using DTT (5 mM final concentration) at 56 °C for 30 min and alkylated with acrylamide (20 mM final concentration) for 30 min at RT, followed by quenching of the reaction through the addition of 5 mM DTT. Finally, trypsin was added at an enzyme-to-sample ratio of 1:50 and the RapiGest concentration adjusted to 0.1% using 0.1 M NH_4HCO_3. Proteins were digested overnight at 37 °C, and on the following day, RapiGest was hydrolyzed by addition of 1% TFA (final concentration) and incubation in a thermomixer at 800 rpm, 37 °C for 30 min. Hydrolyzed RapiGest was precipitated by centrifugation at 20,000× *g*, RT for 10 min and the supernatants were desalted using Oasis HLB cartridges (Waters) as described elsewhere [28]. Briefly, cartridges were equilibrated with 70% ACN, 0.1% acetic acid (AA), washed with 0.1% AA, and the sample was loaded. Subsequently, cartridges were washed with 0.1% AA and peptides were eluted sequentially with 30%, 50%, and 70% ACN, 0.1% AA. Eluate fractions were pooled and the combined samples dried in a vacuum centrifuge. Dried peptides were re-suspended in 5% ACN, and the peptide concentration was determined using the Quantitative Fluorometric Peptide Assay (ThermoFisher Scientific, Waltham, MA, USA), and the peptides were dried again.

2.4. LC-MS/MS Analysis

All analyses were performed using a Dionex Ultimate 3000 nano-UHPLC system coupled to an Orbitrap Fusion Lumos mass spectrometer (both Thermo Fisher Scientific). Analytical columns were produced in-house as follows: spray tips were generated with a P-2000 laser puller (Sutter Instruments, Novato, CA) from 360 µm outer diameter and 100 µm inner diameter fused silica capillaries and packed to a length of 40 cm with 3 µm ReprosilPur AQ C_{18} particles (Dr. Maisch, Ammerbuch-Entringen, Germany). Dried peptides were reconstituted in 5% ACN, 5% formic acid (FA), and 1 µg was loaded together with 750 fmol of internal retention time standards (iRTs, Biognosys, Schlieren, Switzerland) to the analytical column at a flow rate of 600 nl/min with 100% solvent A (0.1% FA in water) for 25 min. Peptides were eluted with 60, 120, and 240 min linear gradients from 5–35% solvent B (90% ACN, 0.1% FA) at a flow rate of 300 nl/min. For parallel reaction monitoring (PRM) measurements, precursor masses were selected from a previously recorded dataset (Table S1, [26]) while the spectral library for data-independent acquisition (DIA) analyses was generated using 240 min data-dependent acquisition (DDA) runs. In these analyses, survey spectra were acquired with a mass range of m/z 350–1200 at a resolution of 60,000 and an AGC target setting of 4×10^5 The most abundant precursor ions (charge states of 2–4) were isolated using the quadrupole (isolation width of m/z 1.6), and fragmented by HCD with a collision energy of 27 in the top speed mode (cycle time of 3 sec). Fragment ion spectra were acquired in the Orbitrap mass analyzer at a resolution of 30,000 and fragmented ions were excluded from further fragmentation for 120 s. For DIA analyses, one MS scan was performed with a mass range of m/z 350–1200, a resolution of 120,000,

a maximum injection time of 20 ms, and an AGC target setting of 5×10^5. The MS scan was followed by static DIA MS/MS scans, covering the same m/z range with an overlap of m/z 0.5, with the following gradient lengths/ scan numbers/ isolation windows/ cycle times: 60 min/ 24 scans/ m/z 35.9/ 2.34 s; 120 min/ 36 scans/ m/z 24.1/ 3.44 s; 240 min/ 58 scans/ m/z 15.2/ 5.45 s). The HCD collision energy was set to 27% and DIA MS/MS scans were acquired in the Orbitrap with a resolution of 30,000, a maximum injection time of 60 ms, and an AGC target setting of 1×10^6. For PRM analyses, MS spectra were acquired with a mass range of m/z 300–1500 at a resolution of 60,000, a maximum injection time of 118 ms, and an AGC target setting of 4×10^5. Peptides were isolated in the quadrupole with an isolation width of m/z 1.2 and fragmented by HCD with a collision energy of 27%. MS/MS scans were acquired in the Orbitrap mass analyzer with a mass range of m/z 200–2000, a resolution of 30,000, a maximum injection time of 54 ms and an AGC target setting of 5×10^4.

2.5. Data Analysis

For DIA library generation, DDA *.raw files were analyzed with the Pulsar search engine integrated in Spectronaut (Version: 14.7.20, Biognosys) (1). Uniprot *Mus musculus* (release date: 09.09.2019 with 17,023 entries), in combination with a database containing common contaminants, was used for database searching with Spectronaut standard settings [29]. In brief, cleavage by trypsin with up to two missed cleavage sites was defined, propionamide (cysteine) was set as fixed and oxidation (methionine) as variable modification, and three to six fragment ions were selected for library generation, dependent on the intensity of the respective peptide. The high-precision iRT concept (dynamic) was applied for retention time alignment. Matching of mass tolerances for precursors, fragment ions, as well as peak extraction windows were determined automatically by Spectronaut. Only MS precursor information was utilized for peak detection, and interference correction was enabled. Global normalization was performed for individual runs based on the median abundance. Data were filtered with a 1% FDR cut off on the precursor and protein level (q-value < 0.01) [30]. *p*-value determination and unsupervised clustering were performed with the post-analysis pipeline of Spectronaut applying default parameters (distance metric: Manhattan Distance; linkage strategy: Ward's method; multiple testing correction: Storey's method).

For PRM analyses, a spectral library was generated using a subset of our previously published DDA dataset [26] with Skyline [31], applying a cut-off score of 0.95. Ambiguous peptide matches were excluded, and the library was filtered for peptides which were previously manually selected to be included in the assay (Table S1). For analysis of PRM data, *.raw files were loaded into Skyline daily version 20.2.1.315. Automated fragment ion selection by Skyline was utilized (6 ions/peptide) with the exception of the peptides with the sequence SLQPLYR and GSFSLSVR, for which only 5 fragment ions matched, using the following criteria: maximum mass error of 10 ppm for MS and MS/MS ion trace filtering (centroid mode) and charge states of 1+/2+ for b- and y-ions as well as 2+/3+ for precursor ions. Integration boundaries of iRT peptides were inspected manually and corrected, if necessary. Experimental data were only reviewed when Skyline reported a peak truncation, and peptides with truncated peaks or no MS/MS signal were excluded from further analysis. Peptide-centric reports were exported and further processed in MS Excel. For peptide and protein quantification, the summed area under the curve (AUC) of fragment ions was used. For all analyses, only peptides with quantitative values in all three replicates were considered.

3. Results and Discussion

We showed previously that the analysis of lysosome-enriched fractions with DIA allows for a superior performance compared to DDA measurements in a reduced amount of time [26]. When we investigated the lysosomal proteome in samples of higher complexity (such as whole cell lysates); however, we observed that the number of lysosomal

proteins that can be reproducibly identified and quantified was markedly lower, indicating a reduced performance in such samples [14]. This is most likely due to the fact that co-fragmenting peptides increase the complexity of MS/MS spectra, which results in a decreased performance for the quantification of lysosomal proteins, as they are of low abundance relative to the whole proteome. In theory, PRM approaches should be superior in this aspect, as only a small m/z window, that is specific for the individual peptide, is fragmented.

In order to determine which strategy is best-suited for the MS/MS-based quantification of the lysosomal proteome in samples with different complexities, we compared DIA- and PRM-based quantification (Figure 1). Initially, we defined a highly reproducible lysosomal proteome from a dataset generated previously by our group, comprising 39 DDA LC-MS/MS runs of lysosome-enriched fractions from MEFs [26]. From these data, we only considered proteins which were assigned to the lysosomal compartment based on gene ontology (GO) and Uniprot categories, and which were detected in $\geq$ 75% of LC-MS/MS runs with $\geq$ 2 unique peptides, resulting in a final list of 374 proteins (Table S1). For the comparison of DIA and PRM, we used a lysosome-enriched fraction from MEFs (LEF) as benchmark samples, as it contains the highest percentage of lysosomal proteins. Furthermore, we used MEF whole cell lysate (MWCL) as well as liver tissue lysate (LTL), representing samples of increasing complexity. We performed all experiments in triplicates with independent experimental replicates for each measurement.

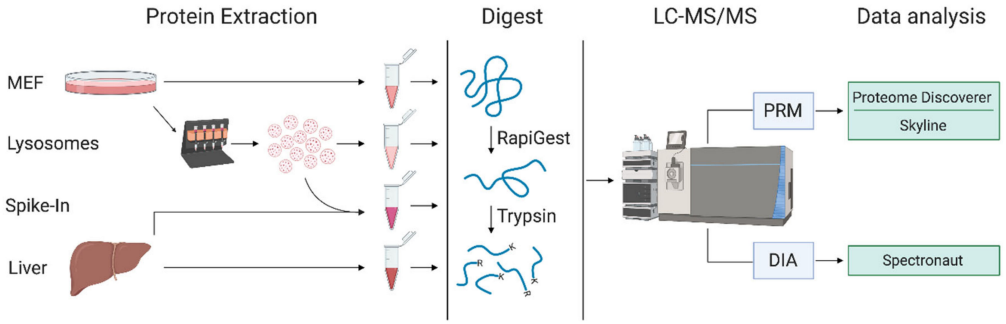


Figure 1. Workflow for sample preparation and analysis. For each sample type, proteins were extracted and digested in three experimental replicates and analyzed by LC-MS/MS using either DIA or PRM with different gradient lengths. MEF: mouse embryonic fibroblasts; DIA: data independent acquisition; PRM: parallel reaction monitoring; K: lysine; R: arginine. Created with Biorender.com.

3.1. Gradient Length and Sample Complexity Affect Lysosomal Protein Quantification by DIA

Initially, we analyzed all three sample types with different gradient lengths (60 min, 120 min, and 240 min) by DIA. We adjusted the width of DIA fragmentation windows depending on the gradient length in order to allow for a similar number of data points across chromatographic peaks of the individual gradients. Consequently, a shorter gradient resulted in a larger m/z window and vice versa, influencing the number of co-fragmented precursor ions. To assess performance of the individual methods, we evaluated both the numbers of total proteins, and those previously reported to be lysosomal (Table S1) that were identified in each run (Table S2).

We found highest total protein numbers in the MWCL, followed by the LEF, and the LTL (Figure 2a). While we observed a steady increase in the number of identified total proteins from 60 min to both 120 min and 240 min gradients for MWCL (increase of 19% and 28%) and LTL (increase of 28% and 46%) samples, the numbers of IDs detected in the LEF only increased from 60 min to 120 min gradients (increase of 30%). When

considering only lysosomal proteins, we identified, as expected, highest numbers in the LEF, followed by MWCL and LTL. While the latter two showed a similar correlation of gradient length and protein identifications, the LEF produced virtually constant numbers for all gradients and only CV values improved. The differences in identifications were particularly pronounced when considering lysosomal proteins quantified with < 5% CV in the 60 min gradient analyses, 116 of which were found in the LEF but only 45 in the LTL (Figure 2b).

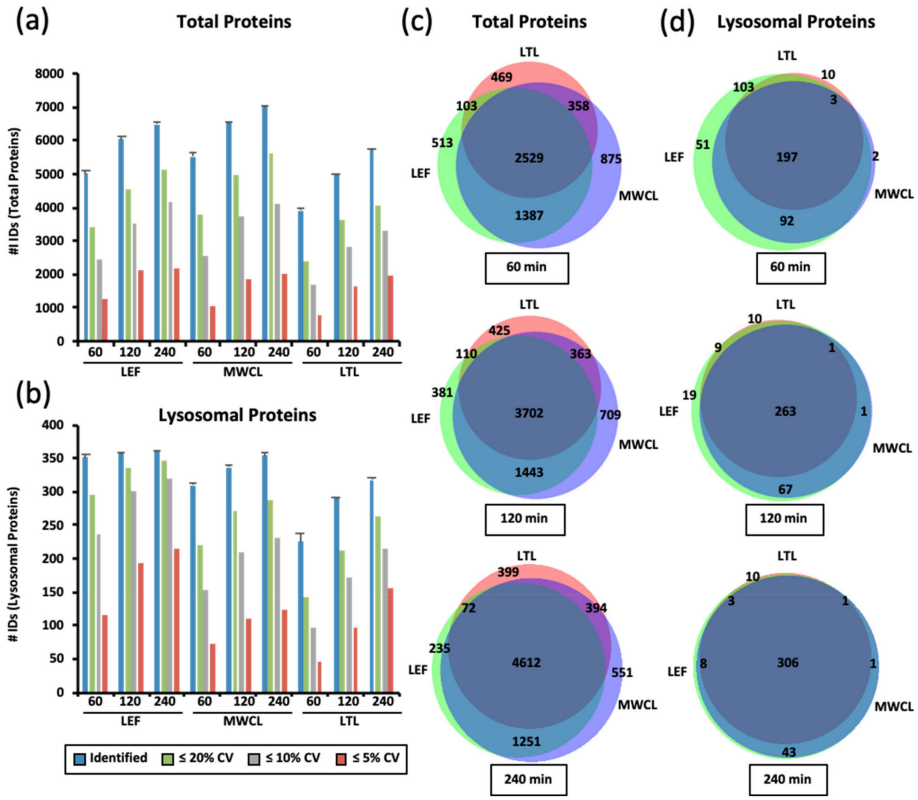


Figure 2. Identification of total and lysosomal proteins in DIA experiments. (**a**,**b**) Numbers of identified total proteins (**a**) and those known to be located at/in the lysosome (**b**). Shown are average values (n = 3) for the analysis of lysosome-enriched fractions of MEF (LEF), MEF whole cell lysate (MWCL), and liver tissue lysate (LTL) with three different gradient lengths. Total IDs as well as those quantified with a CVs $\leq$ 20%, 10%, and 5% are shown. (**c**,**d**) Overlap in protein identification for total proteins (**c**) and those known to be located at the lysosome (**d**) for proteins identified in all three replicates of LEF, MWCL, and LTL samples analyzed with different gradient lengths. Venn diagrams were generated with the tool BioVenn [32]. LEF: lysosome-enriched fractions from mouse embryonic fibroblasts; MWCL: mouse embryonic fibroblast whole cell lysates; LTL: whole liver tissue lysate; CV: coefficient of variation; ID: number of identified proteins.

It was quite surprising to us that we identified the highest number of proteins in the MWCL, as the LTL should theoretically be the most complex sample. A possible explanation for this observation is that the LTL contains a certain number of highly abundant proteins, which account for a larger percentage of the sample than highly abundant MWCL and LEF proteins. Consequently, in LTL the remaining proteins present a smaller fraction of the total

sample. As the C-trap — which is used for ion storage prior to injection into the Orbitrap — has a limited capacity [33], this results in reduced fragment ion intensities for the lower abundant proteins, which are not sufficient for identification/quantification. Furthermore, the highly abundant fragment ions from these proteins dominate the DIA MS/MS spectra, resulting in reduced detection of co-fragmented lower abundant peptides.

When considering the increase in identification of total unique proteins with increasing analysis time, each sample contributed a distinct population (Figure 2c). Lysosomal protein identifications, on the other hand, were very similar between the samples, and for 240 min gradients the majority was identified at least in both the MWCL and the LEF (Figure 2d). These results imply that short gradients suffice to achieve a good coverage of lysosomal proteins in LEFs, while longer gradients are needed when MWCL and LTL are analyzed.

While this confirms that lysosomal proteins are more abundant in lysosome-enriched fractions, which was certainly expected, it also shows that the detection of such lower abundant proteins in DIA analyses suffers from sample complexity. As our DIA analyses were performed with different m/z windows for the different gradient lengths, this effect is most likely related to the number of co-fragmented precursor ions and the resulting MS/MS spectrum complexity. The bigger the fragmentation window is, the more peptides are co-fragmented, and consequently the fragment ions of the lower abundant lysosomal proteins are identified with a lower efficiency.

3.2. Variation of Protein Abundance and Variance between Sample Types in DIA Analyses

To further follow up on this effect, we investigated the protein abundances for the individual samples utilizing median-normalized AUCs (Figure 3a). Confirming our previous assumption, the liver lysate resulted in the highest average abundance (1.6- and 1.8-fold higher compared to the lysosome-enriched fractions and MEF lysate, respectively, for 60 min gradients) and the largest number of highly abundant proteins (36 proteins compared to 11 and 12 with log10 values > 7.5 for LEF and MWCL, respectively). Average protein abundance correlated inversely with the number of protein identifications, with highest values in the shortest gradient, irrespective of the sample type. For lysosomal proteins, we observed highest average abundances in the LEF (1.8- and 2.4-fold higher compared LTL and MWCL for 60 min gradients) and, unlike the total protein identifications, no decrease in abundance with increasing gradient length (Figure 3c). Average CV values, however, behaved similarly for all types of proteins (Figure 3b,d).

To visualize the differences of the individual datasets on a global scale, we generated heatmaps for the average abundances of total and lysosomal proteins, clustered in a row- and column-wise manner (Figure 4a,b). For both analyses, we observed distinct protein populations which formed individual clusters, based on their abundance in the respective sample types and gradient lengths. In most cases, gradient length played a decisive role, while the highest differences existed for the LTL relative to the other samples. For the majority of known lysosomal proteins, we detected a higher abundance in the LEF relative to the MWCL and the LTL, while certain proteins were exclusively identified in the LEF. We also identified, however, some clusters with a higher abundance in MWCL and LTL, implying that either not all lysosomal proteins were recovered efficiently in the lysosome-enrichment step, or that a certain population of these proteins was located in a different cellular compartment.

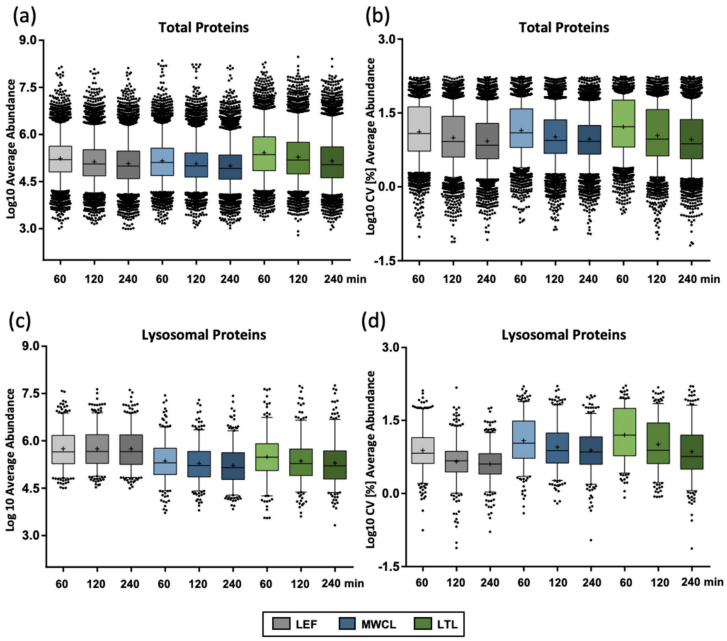


Figure 3. Reproducibility of protein abundance in DIA analyses. (**a**) Protein abundance for total proteins identified in individual samples with different gradient lengths. (**b**) CV values for total proteins identified in individual samples with different gradient lengths. (**c**) Protein abundance for lysosomal proteins identified in individual samples with different gradient lengths. (**d**) CV values for lysosomal proteins identified in individual samples with different gradient lengths. Shown are combined values from 3 replicates, the median is indicated by a line, while the average is marked with a "+". LEF: lysosome-enriched fractions from mouse embryonic fibroblasts; MWCL: mouse embryonic fibroblast whole cell lysates; LTL: whole liver tissue lysate; CV: coefficient of variation.

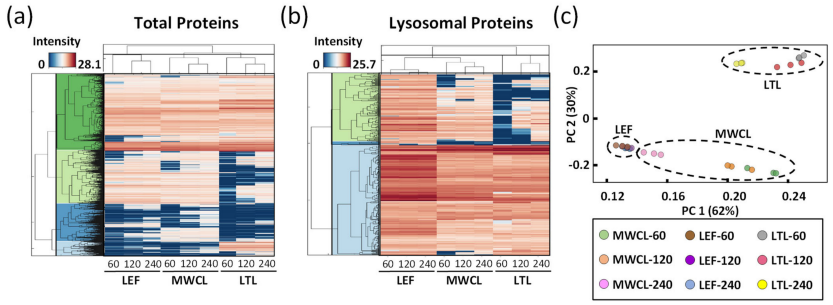


Figure 4. Global analysis of DIA datasets. (**a**,**b**) Unsupervised clustering of average abundances (columns) of LEF, MWCL, and LTL for three different gradient lengths (60, 120 and 240 min) for total proteins identified ((**a**), n = 7145) and lysosomal proteins ((**b**), n = 314). The color code indicates the normalized intensity of the individual proteins. (**c**) Principal component analysis (PCA) for all analyses with two defined standardized principal components (PC1 and PC2). LEF: lysosome-enriched fractions from mouse embryonic fibroblasts; MWCL: mouse embryonic fibroblast whole cell lysates; LTL: whole liver tissue lysate.

Finally, we assessed the global variability between the datasets by principal component analysis (PCA, Figure 4c). The two main principal components (PC1 and PC2), which are responsible for 62% and 30% of the variance in the dataset, allowed for a good separation of the samples. As individual replicates of the same sample type and gradient length clustered closely together, the main variance in the dataset (PC1) can be explained based on the difference of the sample type itself. However, especially for the MWCL, the 240 min gradient data behaved significantly differently than those acquired with other gradients, being actually closer to the LEF. This relates most likely to the fact that the LEF originated from MEFs and that proteome coverage in the 240 min gradient increased to such an extent that very similar proteins were identified (Figure 2c).

3.3. PRM Assay Development

For the 374 proteins included in our lysosomal proteome reference list (Table S1), 10,141 unique peptides were identified in the course of our previous analysis [26]. We narrowed down the list of putative peptides by excluding those identified with variable modifications, missed cleavage sites, or containing the amino acid combination PK or PR (as proline residues interfere with tryptic cleavage). These criteria were fulfilled by 3816 peptides representing 367 proteins. Based on the average signal intensities in this dataset, we considered the two most abundant peptides for each protein, resulting in a final list of 680 peptides, as for some proteins only a single peptide fulfilled our criteria (Figure 5a, Table S1).

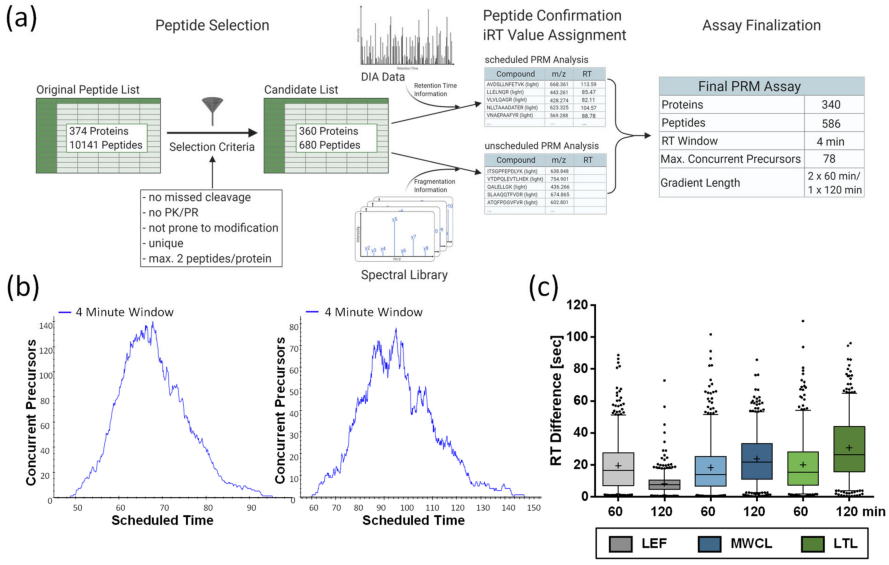


Figure 5. Establishment of PRM assay. (**a**) Workflow for the development of the PRM assay. Created with Biorender.com. (**b**) Distribution of concurrent precursor elution in 60 min and 120 min gradients. (**c**) Differences between predicted and experimentally determined retention times for individual measurements. Shown are combined values of three replicates and the mean (+), median (line), and interquartile range are indicated. LEF: lysosome-enriched fractions from mouse embryonic fibroblasts; MWCL: mouse embryonic fibroblast whole cell lysates; LTL: whole liver tissue lysate; CV: coefficient of variation; RT: retention time.

For PRM assay scheduling, we extracted the peptides' retention time information from the 120 min DIA runs of the LEF, followed by the refinement of the assay by PRM analysis of the same sample, including high-precision indexed retention time (iRT) standards [29].

After analysis with an initial scheduled PRM assay (15 min retention time windows), we performed unscheduled PRM runs for those peptides that were not detected in these initial analyses. Finally, we combined the acquired retention times of all peptides that we were able to detect with distinct fragment ion signatures, and created an iRT-normalized library. This resulted in a final assay comprising 586 peptides from 340 lysosomal proteins. For the analysis of acquired PRM data with Skyline [31], we built a reference spectral library from our previously measured DDA dataset of the LEF [26]. Finally, we generated two assays utilizing 4 min retention time windows. In one assay, all peptides were analyzed in a single 120 min gradient, while the other assay consisted of two 60 min gradients. This was necessary, as the high number of concurrent precursors (up to 140) would have drastically reduced the number of data points for chromatographic peaks eluting in the middle of gradient (Figure 5b). For subsequent analyses, we determined iRT correction factors for the different sample types using DDA runs, and adjusted the scheduling accordingly. For data export from Skyline, a minimum number of six data points was defined.

3.4. Gradient Length and Sample Type Affect Data Quality in PRM Analyses

Due to the high number of peptides, we defined parameters for the acceptance of PRM quantification data without manual inspection of each peptide. Initially, we assessed the difference between predicted and experimentally observed retention times. For the 60 min and 120 min gradients, we observed average peak widths of 21 sec and 25 sec, respectively, and an average retention time variability of $\pm$15 sec, with a slightly lower average shift for the 60 min gradients (Figure 5c). LEF analyses with 120 min gradients presented with only $\pm$8 sec an exception, which might be due to the fact that we performed the PRM assay retention time normalization with 120 min gradient measurements of the LEF, while the scheduling for MWCL and LTL was solely based on iRT predictions.

Next, we investigated the quality of acquired fragment ions for the individual peptides utilizing the dot product (dotP) value [34], which allows for correlation between the acquired spectrum and the spectral library (generated from our reference dataset [26]). Especially for the analysis of unfractionated highly complex samples, this allows to identify the impact of interfering ions that may result in false quantification results. Across all analyses, the average dotP value was > 0.85, indicating a good matching of our PRM data with the spectral library (Table S3). While we observed roughly similar dotP values for both the 60 min and 120 min assays, they decreased with sample complexity (Figure 6a). Compared to the LEF, which displayed the least variation, especially the LTL resulted in lower dotP values and higher variability. These findings imply a lower relative abundance of lysosomal proteins and an increase of interfering fragment ions in the MS/MS spectra for the more complex samples, which is also in agreement with the DIA data.

Subsequently, we investigated the correlation of dotP values and numbers of fragment ions used for quantification of the different sample types and gradient lengths (Figure S1). Utilizing three to six fragment ions, we applied different dotP value thresholds and determined the number of peptides passing it. As expected, lower numbers of fragment ions resulted in more peptides passing the threshold at higher dotP values. This was especially true when dotP thresholds $\geq$ 0.9 were applied, as we observed a clear difference between the peptides identified with 3, 4, 5, or 6 fragment ions. For lower dotP values (0.7–0.8), this effect was far less pronounced. As already indicated by the average dotP values (Figure 6a), an inverse correlation with sample complexity could be observed. Based on these analyses, we defined 6 fragment ions per peptide with a dotP value of 0.7 as cut-off for the acceptance of quantification information from PRM data.

For LEF data, this cut-off resulted in an acceptance rate of 92% of the peptides included in our assay for both gradients. The value for MWCL was 87% for both gradients and for LTL 73% and 78%, for the 60 min and 120 min gradient, respectively. Applying these cut-offs, we exported the data from Skyline and utilized them for all further analyses (Table S4).

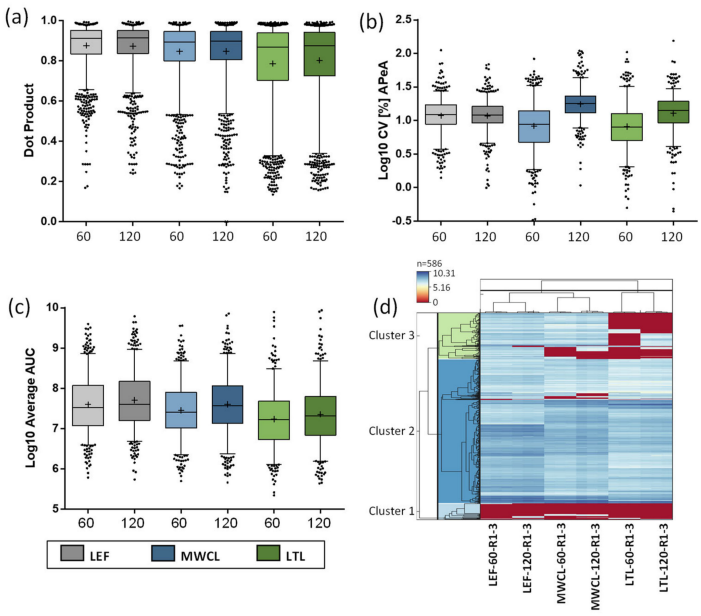


Figure 6. Characterization of data generated by PRM analysis. (**a**) Assessment of data quality in PRM analyses by Dot Product values. (**b**) Log10 CV values for the summed AUC of all targeted peptides. (**c**) Averaged log10-transformed summed peptide AUCs of the replicates (n = 3) across the different sample types and gradient lengths from PRM measurements. (a/b/c) Indicated are mean (+), median (line), and interquartile range. (**d**) Heatmap of the log10-transformed AUCs of all peptides covered by the PRM assay across all measurements. Each column contains data from one measurement and each row represents one peptide. Peptides are clustered if they exhibit similar trends across the samples. LEF: lysosome-enriched fractions from mouse embryonic fibroblasts; MWCL: mouse embryonic fibroblast whole cell lysates; LTL: whole liver tissue lysate; AUC: area under the curve; CV: coefficient of variation; APeA: average peptide abundance.

3.5. PRM Analysis of the Lysosomal Proteome

We initially investigated the reproducibility of quantification (Figure 6b). While CVs of the LEF analyses were similar for both gradients, the 120 min gradient resulted in consistently higher CVs for both MWCL and LTL. Surprisingly, the higher complexity samples resulted in a lower average CV than the LEF for the 60 min runs. A possible explanation for this observation is the lower sample amount utilized for LEF sample preparation (~20 µg) compared to MWCL and LTL (~100 µg), which may have resulted in a higher variability during pipetting and desalting. Subsequently, we calculated the average summed area under the curve (AUC) for each sample type and gradient length (Figure 6c). Interestingly, while we saw a higher summed abundance for the LEF in comparison to the other samples for 60 min gradients, the values were more similar for the 120 min analyses, especially for the comparison of LEF and MWCL. This could be related to the different numbers of data points acquired over the chromatographic peak as well as variances in peak width/shape between gradients.

Finally, we assessed the overall correlation of the data in a heatmap, depicting the signal intensities of individual peptides in each sample and replicate, clustered in a row- and column-wise manner (Figure 6d). In general, the lysosomal peptides formed three distinct clusters. Two clusters showed similar expression levels (general high or low expression) in all samples, while the third cluster contained proteins that were detected

with differing levels in the individual samples. In agreement with the DIA data, we observed subsets of peptides that were only found in the LEF, and were not detected in both the MWCL and LTL. Moreover, peptides existed that were detected both in LEF and MWCL, but not in LTL, and a small subset with higher abundance in LTL compared to the other samples.

3.6. Comparison of DIA- and PRM-Based Quantification of the Lysosomal Proteome

In order to correlate the performance of DIA- and PRM-based quantification of the lysosomal proteome, we initially compared the data obtained from the individual datasets. When considering the average abundance and CVs of lysosomal proteins, we observed for all DIA analyses that longer gradients resulted in lower CV values but also lower abundances (Figure 3c,d), while for PRM both intensities and CVs (with the exception of the LEF data) increased with gradient length (Figure 6b,c). When considering how many lysosomal proteins were found in the individual analyses, we identified higher numbers for DIA, with the exception of LTL analyzed with 60 min gradients (Figure S2). To assess to what extend the abundances acquired with the individual approaches correlate, we extracted the AUCs of the peptides included in our PRM assay from the DIA dataset (60 min gradients for both approaches) and performed a direct comparison (Figure 7a). We observed for all three sample types that PRM resulted in higher signal intensities than DIA and that correlation of signal intensities was dependent on the abundance of the respective protein. We observed a good correlation for high-abundant proteins (upper 50% of DIA intensities) in all sample types. For low-abundant proteins (lower 50% of DIA intensities), we only detected a good correlation between DIA and PRM for the LEF. For the more complex samples, however, DIA seemed to underestimate high signal intensities, resulting in poor correlation with the PRM data.

As the main application of both methods is the quantitative comparison of the lysosomal proteome between different states, we performed a spike-in experiment to simulate constitutive upregulation of the whole lysosomal proteome and analyzed the sample by both PRM and DIA, applying 60 min gradients for both approaches. For this purpose, we combined LEF and LTL in a 1 to 5 ratio and compared the data to LTL samples without spike-in (Figure 1). In theory, as LEFs contain higher amounts of lysosomal proteins, this should result in a general increase of intensity for all lysosomal proteins present in the sample.

For both approaches, the number of detected lysosomal proteins increased in comparison to LTL without spike-in, while the increase for PRM was 50% higher compared to DIA (223 to 243 for DIA and 278 to 308 for PRM, Figure 7b). Subsequently, we investigated the fold change ratios for proteins identified in all samples with both approaches (Figure 7c). We detected a median increase of intensity of 1.8 for PRM and 1.2 for DIA. When investigating values for individual proteins, we observed a discrepancy of $\geq 30\%$ between fold change values acquired by DIA and PRM for 75% of proteins (average CV for DIA and PRM analysis of LTLs: 16% and 7%). While 142 proteins were detected with a higher value in PRM, only 35 were higher in the DIA data (Figure 7d). Classification of proteins based on their fold change values between the spike-in and the LTL sample further showed that DIA failed to detect any increase in signal intensity for 81 proteins upon spike-in, while this was only the case for 10 proteins in the PRM data (Figure 7e). Subsequently, we investigated if this effect was related to the abundance of individual proteins, as we observed markedly reduced correlation coefficients between DIA and PRM for lower abundant proteins in LTL samples (Figure 7a). Along this line, we grouped all proteins based on their abundance relative to the highest/lowest abundant protein in the respective dataset and plotted the observed fold change ratios for the individual groups (Figure 7f). While we observed highly similar fold change value distributions between LTL and spike-in samples for proteins across the whole range of abundance for the PRM data, a clear shift in the pattern of the DIA data was visible. Relative to the PRM data, DIA reported higher fold change ratios for low-abundant proteins while it resulted in lower

values for high-abundant proteins. Taken together, these data indicate that PRM is better suited for the quantification of changes in the lysosomal proteome of LTL, which is mainly related to the better performance for the highest- and lowest-abundant lysosomal proteins in the dataset.

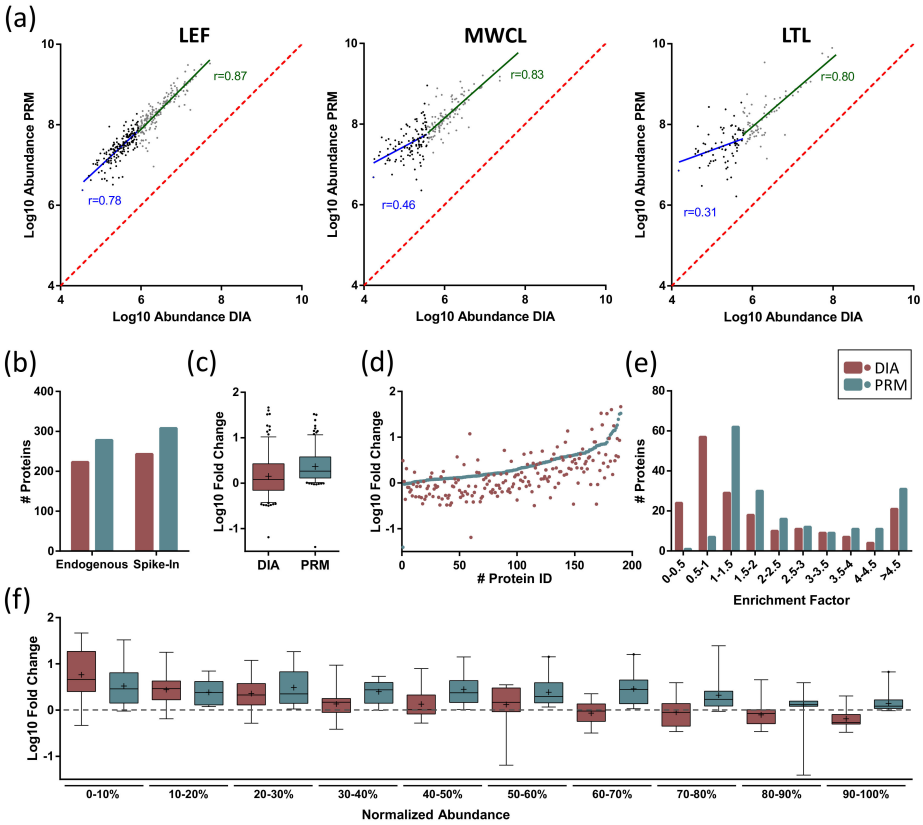


Figure 7. Comparison of DIA and PRM analysis of lysosomal proteins. (**a**) Correlation of normalized peptide signal intensities for DIA and PRM runs. Proteins are sorted based on their intensity in DIA measurements and grouped into two groups based on their intensity (upper/lower 50% of proteins). For each population, a linear regression analysis was performed and the respective correlation coefficient (r) is indicated. (**b**) Identification of proteins in LTL with and without spike-in of LEFs. (**c**,**d**) Fold change values for individual lysosomal proteins in LTL with spike-in of LEFs. (**e**) Frequency of proteins within distinct fold change quantiles for DIA and PRM data for ratios of LTL with/without spike-in of LEFs. (**f**) Protein fold change values for LTL with/without spike-in of LEFs. Proteins are grouped based on their abundance in the respective dataset relative to the highest/lowest-abundant protein. LEF: lysosome-enriched fractions from mouse embryonic fibroblasts; MWCL: mouse embryonic fibroblast whole cell lysates; LTL: whole liver tissue lysate.

4. Conclusions

In the present study, we analyzed the lysosomal proteome in samples of varying complexity by DIA and PRM. While both methods were well-suited for the analysis of lysosomal proteins in all samples, differences between the approaches became apparent that were mostly related to sample complexity. DIA identified more proteins in lower complexity samples and at longer gradients, since it was not limited by a predefined list of

peptides, as was the case for the PRM analyses. Furthermore, no assay development was necessary for DIA analyses, thus greatly reducing the amount of time needed. For peptides covered by both approaches, DIA and PRM performed similarly for lower complexity LEFs, while PRM outperformed DIA in both MWCL and LTL. Especially for the quantification of protein level changes in LTL, PRM was able to identify significantly higher numbers of protein level alterations than DIA, which reported no change in abundance for a high number of proteins. Therefore, for the analysis of highly complex samples, such as whole tissue lysates, PRM presents the method of choice. Our developed PRM assay allows for the direct analysis of the lysosomal proteome from small amounts of whole tissue samples, without the need for lysosome enrichment, extending the toolbox for the investigation of the lysosomal proteome in complex samples.

Supplementary Materials: The following are available online at https://www.mdpi.com/2227-7382/9/1/4/s1, Table S1: Reference list of high confidence lysosomal proteins and peptides covered by PRM assay. Supplementary Table S2: DIA Data, direct output from Spectronaut as well as processed data. Supplementary Table S3: PRM data, direct output from Skyline. Supplementary Table S4: PRM data filtered for dot p values > 0.7 and further analyses. Supplementary Figure S1: Dot product threshold determination for the acceptance of PRM data. Figure S2: Overlap of identified proteins from PRM and DIA runs.

Author Contributions: Conceptualization, D.W.; methodology, P.M., A.S., and J.S.; validation, P.M., A.S., and J.S.; formal analysis, P.M., A.S., and J.S.; investigation, P.M., A.S., and J.S.; resources, P.M., and A.S.; writing—original draft preparation, A.S., and D.W.; writing—review and editing, P.M., A.S., J.S., and D.W.; visualization, P.M., A.S., and J.S.; supervision, D.W.; project administration, D.W.; funding acquisition, D.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deutsche Forschungsgemeinschaft, grant number FOR2625, and the Studienstiftung des Deutschen Volkes.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [35] partner repository with the dataset identifier PXD023278.

Acknowledgments: The authors would like to thank Robert Hardt for helpful discussions, Shiva Ahmadi for help with liver samples, and Fatema Akter as well as Srigayatri Ponnaiyan for help with lysosome-enriched fractions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. De Duve, C.; Wattiaux, R. Functions of Lysosomes. *Annu. Rev. Physiol.* **1966**, *28*, 435–492. [CrossRef] [PubMed]
2. Perera, R.M.; Zoncu, R. The Lysosome as A Regulatory Hub. *Annu. Rev. Cell Dev. Biol.* **2016**, *32*, 223–253. [CrossRef] [PubMed]
3. Platt, F.M.; d'Azzo, A.; Davidson, B.L.; Neufeld, E.F.; Tifft, C.J. Lysosomal Storage Diseases. *Nat. Rev. Dis. Primers* **2018**, *4*, 27. [CrossRef] [PubMed]
4. Ballabio, A.; Gieselmann, V. Lysosomal Disorders: From Storage to Cellular Damage. *Biochim. Biophys. Acta* **2009**, *1793*, 684–696. [CrossRef] [PubMed]
5. Jeyakumar, M.; Dwek, R.A.; Butters, T.D.; Platt, F.M. Storage Solutions: Treating Lysosomal Disorders of The Brain. *Nat. Rev. Neurosci.* **2005**, *6*, 713–725. [CrossRef] [PubMed]
6. Perera, R.M.; Stoykova, S.; Nicolay, B.N.; Ross, K.N.; Fitamant, J.; Boukhali, M.; Lengrand, J.; Deshpande, V.; Selig, M.K.; Ferrone, C.R.; et al. Transcriptional Control of Autophagy-Lysosome Function Drives Pancreatic Cancer Metabolism. *Nature* **2015**, *524*, 361–365. [CrossRef]
7. DePaolo, J.; Goker-Alpan, O.; Samaddar, T.; Lopez, G.; Sidransky, E. The Association Between Mutations in The Lysosomal Protein Glucocerebrosidase And Parkinsonism. *Mov. Disord.* **2009**, *24*, 1571–1578. [CrossRef]
8. Chi, C.; Riching, A.S.; Song, A.K. Lysosomal Abnormalities in Cardiovascular Disease. *Int. J. Mol. Sci.* **2020**, *21*, 811. [CrossRef]
9. Ballabio, A.; Bonifacino, J.S. Lysosomes as Dynamic Regulators of Cell and Organismal Homeostasis. *Nat. Rev. Mol. Cell Biol.* **2019**, *2019*, 1–18. [CrossRef]

10. Settembre, C.; Di Malta, C.; Polito, V.A.; Arencibia, M.G.; Vetrini, F.; Erdin, S.; Erdin, S.U.; Huynh, T.; Medina, D.; Colella, P.; et al. Tfeb Links Autophagy to Lysosomal Biogenesis. *Science* **2011**, *332*, 1429–1433. [CrossRef]
11. Zhang, C.S.; Jiang, B.; Li, M.; Zhu, M.; Peng, Y.; Zhang, Y.L.; Wu, Y.Q.; Li, T.Y.; Liang, Y.; Lu, Z.; et al. The Lysosomal V-Atpase-Ragulator Complex Is A Common Activator for Ampk And Mtorc1, Acting as A Switch Between Catabolism and Anabolism. *Cell Metab.* **2014**, *20*, 526–540. [CrossRef] [PubMed]
12. Aebersold, R.; Mann, M. Mass-Spectrometric Exploration of Proteome Structure and Function. *Nature* **2016**, *537*, 347–355. [CrossRef] [PubMed]
13. Akter, F.; Ponnaiyan, S.; Koegler-Mohrbacher, B.; Bleibaum, F.; Damme, M.; Renard, B.Y.; Winter, D. Multi Cell Line Analysis of Lysosomal Proteomes Reveals Unique Features and Novel Lysosomal Proteins. *bioRxiv* **2020**. [CrossRef]
14. Singh, J.; Kaade, E.; Muntel, J.; Bruderer, R.; Reiter, L.; Thelen, M.; Winter, D. Systematic Comparison of Strategies for The Enrichment of Lysosomes by Data Independent Acquisition. *J. Proteome Res.* **2020**, *19*, 371–381. [CrossRef] [PubMed]
15. Thelen, M.; Winter, D.; Braulke, T.; Gieselmann, V. Silac-based comparative proteomic analysis of lysosomes from mammalian cells using Lc-Ms/Ms. In *Lysosomes: Methods and Protocols*; Öllinger, K., Appelqvist, H., Eds.; Springer: New York, NY, USA, 2017; pp. 1–18.
16. Wyant, G.A.; Abu-Remaileh, M.; Frenkel, E.M.; Laqtom, N.N.; Dharamdasani, V.; Lewis, C.A.; Chan, S.H.; Heinze, I.; Ori, A.; Sabatini, D.M. Nufip1 Is A Ribosome Receptor for Starvation-Induced Ribophagy. *Science* **2018**, *360*, 751–758. [CrossRef] [PubMed]
17. Zoncu, R.; Bar-Peled, L.; Efeyan, A.; Wang, S.; Sancak, Y.; Sabatini, D.M. Mtorc1 Senses Lysosomal Amino Acids Through an Inside-Out Mechanism That Requires the Vacuolar H+-Atpase. *Science* **2011**, *334*, 678–683. [CrossRef] [PubMed]
18. Shi, T.; Song, E.; Nie, S.; Rodland, K.D.; Liu, T.; Qian, W.J.; Smith, R.D. Advances in Targeted Proteomics and Applications To Biomedical Research. *Proteomics* **2016**, *16*, 2160–2182. [CrossRef]
19. Bilbao, A.; Varesio, E.; Luban, J.; Strambio-De-Castillia, C.; Hopfgartner, G.; Müller, M.; Lisacek, F. Processing Strategies and Software Solutions for Data-Independent Acquisition in Mass Spectrometry. *Proteomics* **2015**, *15*, 964–980. [CrossRef]
20. Hoofnagle, A.N.; Becker, J.O.; Oda, M.N.; Cavigiolio, G.; Mayer, P.; Vaisar, T. Multiple-Reaction Monitoring-Mass Spectrometric Assays Can Accurately Measure the Relative Protein Abundance in Complex Mixtures. *Clin. Chem.* **2012**, *58*, 777–781. [CrossRef]
21. Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. Selected Reaction Monitoring for Quantitative Proteomics: A Tutorial. *Mol. Syst. Biol.* **2008**, *4*, 222. [CrossRef]
22. Schmidt, A.; Gehlenborg, N.; Bodenmiller, B.; Mueller, L.N.; Campbell, D.; Mueller, M.; Aebersold, R.; Domon, B. An Integrated, Directed Mass Spectrometric Approach for In-Depth Characterization of Complex Peptide Mixtures. *Mol. Cell. Proteom.* **2008**, *7*, 2138–2150. [CrossRef] [PubMed]
23. Damaghi, M.; Tafreshi, N.K.; Lloyd, M.C.; Sprung, R.; Estrella, V.; Wojtkowiak, J.W.; Morse, D.L.; Koomen, J.M.; Bui, M.M.; Gatenby, R.A.; et al. Chronic Acidosis in The Tumour Microenvironment Selects for Overexpression of Lamp2 in the Plasma Membrane. *Nat. Commun.* **2015**, *6*, 8752. [CrossRef] [PubMed]
24. Sjödin, S.; Öhrfelt, A.; Brinkmalm, G.; Zetterberg, H.; Blennow, K.; Brinkmalm, A. Targeting Lamp2 In Human Cerebrospinal Fluid with A Combination of Immunopurification and High Resolution Parallel Reaction Monitoring Mass Spectrometry. *Clin. Proteom.* **2016**, *13*, 4. [CrossRef] [PubMed]
25. Ivry, S.L.; Knudsen, G.M.; Caiazza, F.; Sharib, J.M.; Jaradeh, K.; Ravalin, M.; O'Donoghue, A.J.; Kirkwood, K.S.; Craik, C.S. The Lysosomal Aminopeptidase Tripeptidyl Peptidase 1 Displays Increased Activity in Malignant Pancreatic Cysts. *Biol. Chem.* **2019**, *400*, 1629–1638. [CrossRef] [PubMed]
26. Ponnaiyan, S.; Akter, F.; Singh, J.; Winter, D. Comprehensive Draft of The Mouse Embryonic Fibroblast Lysosomal Proteome by Mass Spectrometry Based Proteomics. *Sci. Data* **2020**, *7*, 1. [CrossRef]
27. Muntel, J.; Gandhi, T.; Verbeke, L.; Bernhardt, O.M.; Treiber, T.; Bruderer, R.; Reiter, L. Surpassing 10 000 Identified and Quantified Proteins in A Single Run by Optimizing Current Lc-Ms Instrumentation and Data Analysis Strategy. *Mol. Omics* **2019**, *15*, 348–360. [CrossRef]
28. Winter, D.; Steen, H. Optimization of Cell Lysis and Protein Digestion Protocols for The Analysis of Hela S3 Cells by Lc-Ms/Ms. *Proteomics* **2011**, *11*, 4726–4730. [CrossRef]
29. Bruderer, R.; Bernhardt, O.M.; Gandhi, T.; Reiter, L. High-Precision Irt Prediction in The Targeted Analysis of Data-Independent Acquisition and Its Impact on Identification and Quantitation. *Proteomics* **2016**, *16*, 2246–2256. [CrossRef]
30. Rosenberger, G.; Bludau, I.; Schmitt, U.; Heusel, M.; Hunter, C.L.; Liu, Y.; MacCoss, M.J.; MacLean, B.X.; Nesvizhskii, A.I.; Pedrioli, P.G.; et al. Statistical Control of Peptide and Protein Error Rates in Large-Scale Targeted Data-Independent Acquisition Analyses. *Nat. Methods* **2017**, *14*, 921–927. [CrossRef]
31. Pino, L.K.; Searle, B.C.; Bollinger, J.G.; Nunn, B.; MacLean, B.; MacCoss, M.J. The Skyline Ecosystem: Informatics for Quantitative Mass Spectrometry Proteomics. *Mass Spectrom. Rev.* **2020**, *39*, 229–244. [CrossRef]
32. Hulsen, T.; de Vlieg, J.; Alkema, W. Biovenn—A Web Application for The Comparison and Visualization of Biological Lists Using Area-Proportional Venn Diagrams. *BMC Genom.* **2008**, *9*, 488. [CrossRef] [PubMed]
33. Zubarev, R.A.; Makarov, A. Orbitrap Mass Spectrometry. *Anal. Chem.* **2013**, *85*, 5288–5296. [CrossRef] [PubMed]

34. Frewen, B.E.; Merrihew, G.E.; Wu, C.C.; Noble, W.S.; MacCoss, M.J. Analysis of Peptide Ms/Ms Spectra from Large-Scale Proteomics Experiments Using Spectrum Libraries. *Anal. Chem.* **2006**, *78*, 5678–5684. [CrossRef] [PubMed]
35. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The Pride Database and Related Tools and Resources In 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47*, D442–D450. [CrossRef] [PubMed]

4.3 Publication: Supplementary Information

Supplementary Material

Targeted Quantification of the Lysosomal Proteome in Complex Samples

Peter Mosen*, Anne Sanner*, Jasjot Singh, and Dominic Winter#

Institute for Biochemistry and Molecular Biology,
Medical Faculty, University of Bonn, 53115 Bonn, Germany

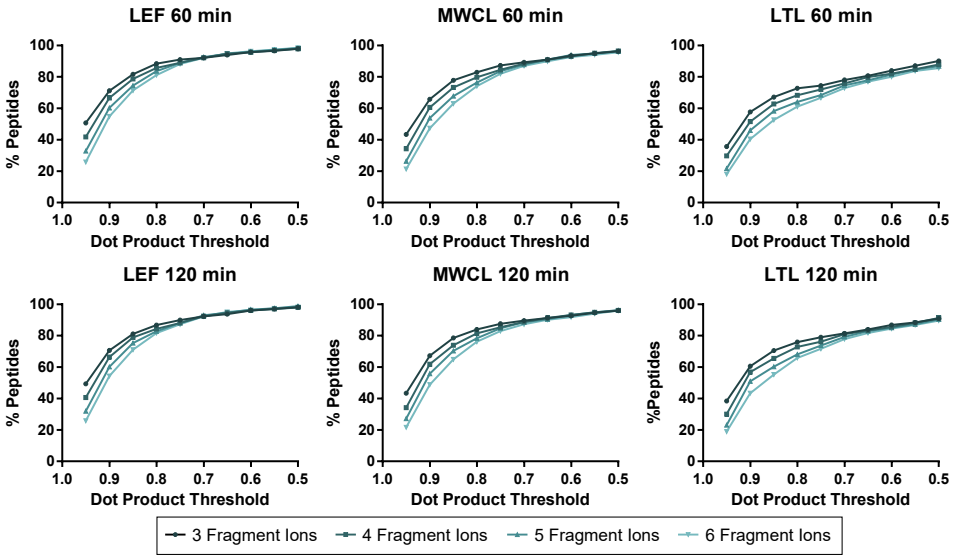
*these authors contributed equally

#Corresponding author
Dr. Dominic Winter
Institute for Biochemistry and Molecular Biology
Medical Faculty
Rheinische Friedrich-Wilhelms-University of Bonn
Nussallee 11, 53115 Bonn, Germany
Email: dominic.winter@uni-bonn.de
Tel: +49 228 73 7081

Supplementary Figure 1: Dot product threshold determination for the acceptance of PRM data.
Supplementary Figure 2: Overlap of identified proteins from PRM and DIA runs

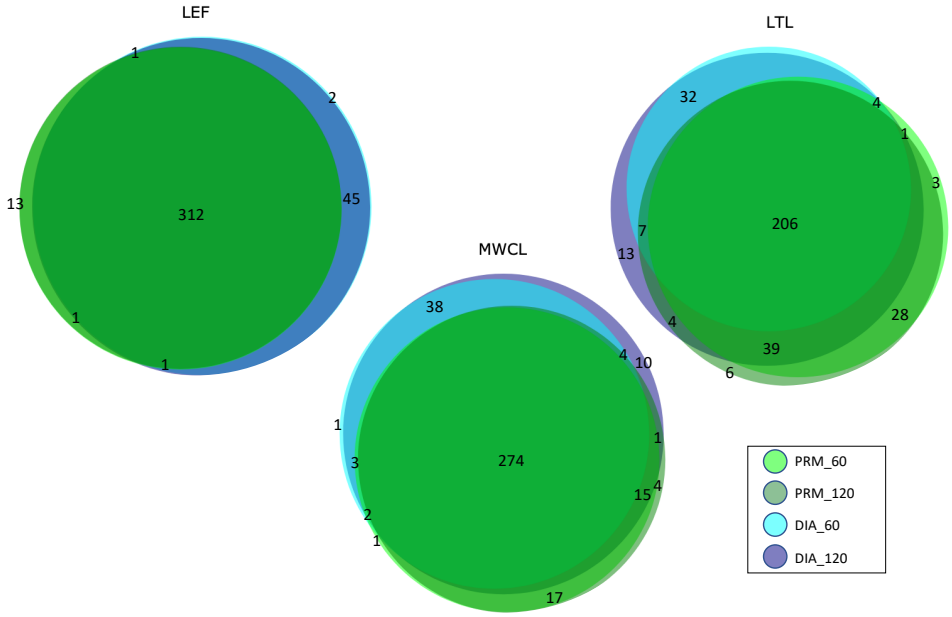
Supplementary Reference

1. Hulsen, T., J. de Vlieg, and W. Alkema, BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. Bmc Genomics, 2008. 9(488).



Supplementary Figure 1: Dot product threshold determination for the acceptance of PRM data. Shown are threshold determination graphs for 60 min and 120 min gradients for the lysosome-enriched fraction (LEF), mouse embryonic fibroblast whole cell lysate (MWCL), and liver tissue lysate (LTL). Data were analyzed with Skyline Daily using three, four, five, or six fragment ions. In each plot, the percentage of peptides passing the applied dot product threshold, which is indicated on the X axis, is displayed. Each line represents a different number of fragment ions utilized for peptide quantification.

S2



Supplementary Figure 2: Overlap of identified proteins from PRM and DIA runs. PRM data were analyzed using Skyline Daily, DIA data using Spectronaut. Proteins which were quantified in individual MS runs were extracted and compared. Each Venn diagram represents one sample type: lysosome-enriched fraction (LEF), mouse embryonic fibroblast whole cell lysate (MWCL), and liver tissue lysate (LTL). Compared are data from four different triplicate analyses conducted with PRM and DIA, each with 60 min and 120 min gradients. Venn diagrams were created using DeepVenn [1].

S3

4.4 Conclusion

Different gradient length and associated with this also adjusted MS parameters ensuring study comparability, were used. The quantitative capabilities of the two data acquisition strategies were evaluated in samples of varying complexities, ranging from isolated lysosomes of low complexity (LEF) to moderately complex **m**ouse embryonic fibroblast (MEF) **w**hole **c**ell **l**ysates (MWCL) and highly complex liver tissue lysates (LTL). With its unbiased, comprehensive data acquisition strategy, DIA-MS was able to be applied immediately, beyond that not only allowing for quantification of lysosomal but also non-lysosomal proteins. For the analysis of lysosomes, acquired results using DIA imply that for low complex samples such as LEF short gradients of even 60 minutes are absolutely sufficient to achieve highest lysosomal proteome coverage. In case of medium complex (MWCL) and high complex samples (LTL) however, longer gradients for MS data acquisition are necessary and advisable to maximize lysosomal proteome coverage. Of note, as the gradient length increases there is an enhancement in the coefficient of variation (CV) for protein quantification (Figure 2B) using DIA. In general, data indicated that DIA suffers in the detection of lower abundant (lysosomal) proteins with increasing sample complexity, resulting in lowest identification numbers for total proteins (Figure 2A) and lysosomal proteins (Figure 2B) in the liver tissue lysate sample (LTL), even though the liver tissue lysate sample gives rise to the highest (total proteins, Figure 3A) and second highest (lysosomal proteins, Figure 3C) average abundances. Most likely this phenomenon is a combined effect of sample complexity itself, where a set of extremely high abundant proteins accounts for a significant proportion of the proteome (compared to the remaining lower abundant proteins) and the operational mode of OT-based instruments, disadvantaging the detection of low-abundant precursor- and fragment ions [256].

In such a scenario of sample complexity, targeted data acquisition as performed with PRM is expected to be superior. The narrow precursor isolation around the predefined peptide precursor m/z (typically around m/z 1, as opposed to the m/z 35.9, 24.1 and 15.2 wide window isolation in the DIA methods) minimizes precursor co-isolation and drastically limits co-fragmentation of e.g. high abundant proteins which might have dominated in DIA the MS2 spectrum, further boosting signal intensities of targeted, low abundant (lysosomal) protein. However, PRM acquisition requires manual effort in assay development/optimization and is constrained to a predefined set of peptides. Based on previously generated data in the group [241] and a highly reproducible selection of lysosomal annotated proteins (found in more than 75% of MS-experiments and represented with more than or equals to 2 unique peptides), a PRM assay comprising 340 lysosomal proteins with 586 peptides, was developed in a multistep process (Figure 5A). Data acquisition was performed in either one long gradient (120 minutes) or two short gradients of each 60 minutes (Figure 5B) covering the full set of 586 peptides. For semi-manual data analysis and processing, a spectral library and retention time library were built, threshold values for user-based quantification refinement were determined (Figure

S1) and a workflow for data processing was established. Similar to the DIA experiments, also the quantification of lysosomal peptides/proteins faces challenges with increasing sample complexity, as reflected by decreasing dotP value (a measure for the degree of similarity between reference and experimental MS2 spectrum), which is related to the increasing MS2 complexity (Figure 6A). Unlike the DIA experiments, the application of shorter gradients (60 minutes) results in lower, i.e. better, CVs for lysosomal peptide quantification (Figure 6B).

In the direct comparison between the PRM approach and DIA strategy applied, DIA, with its unbiased data acquisition mode, was able to quantify a higher number of lysosomal proteins in combination with longer gradients in low (LEF) and medium (MWCL) complex samples (Figure S2). In high complex samples (LTL) however, PRM detects more lysosomal proteins (Figure 7A, left). Furthermore, PRM resulted in clearly higher signal intensities in all three sample types (Figure 7A), and while a good correlation for peptides of high abundant proteins across all sample types was observed (upper subpopulation) for the peptides of low abundant proteins (lower subpopulation) correlation between DIA and PRM data decreases drastically with increasing complexity (Figure 7A). This suggests that DIA with increasing sample complexity tends to underestimate these signal intensities and that the quantification of subtle changes might not be reflected correctly. To test this hypothesis a spike-in experiment, adding a defined amount of the enriched lysosome fraction to complex liver tissue lysate (LTL) was performed, thereby simulating the upregulation of lysosomal proteins. The comparison of protein quantities with and without LEF spike-in allowed a statement, how similar (and well) the respective MS approach (DIA or PRM) reflects the quantitative change (upregulation) in protein abundance. Here, for the detection and quantification of subtle changes in protein quantities from complex samples, PRM outperforms DIA. PRM was able to detect clearly more lysosomal proteins with a protein (up)regulation (Figure 7A and 7D). For the majority of proteins (75%) a considerable discrepancy (≥ 30%) in the quantified fold-changes between PRM and DIA was found (Figure 7D), whereas DIA quantification failed to detect any change in protein quantity, especially for subtle (up)regulation, and reporting no or a negative regulation, for approximately 40% of proteins (Figure 7E). Given these results, and with regard to the future study of LSDs from tissue samples, a targeted MS acquisition strategy, like PRM or MRM, emerges as the preferred approach for accurate (absolute) quantification of low-abundance lysosomal proteins from whole cell/tissue lysates.

5. Chapter 3 - Targeted Absolute Quantification of Lysosomal Proteins using QconCAT Peptide Standards and MRM Mass Spectrometry

Remark: The results presented in this chapter have been published as a preprint on bioRxiv and have not yet undergone formal peer review.

Based on: Mosen, P.R., et al., *Absolute Quantification of Lysosomal Proteins by Multiple Reaction Monitoring Mass Spectrometry and QconCAT Protein Standards*. bioRxiv, 2025. 01.09.632238.; accessible via: https://doi.org/10.1101/2025.01.09.632238

5.1 Introduction

The view on lysosomes as static and unregulated organelles, which prevailed for decades, is currently shifting, recognizing their highly dynamic and motile nature [19, 62, 106, 257] along with the diverse roles and functions in the cell [10, 11, 13-18, 258]. In this context, comparative studies utilizing MS have expanded our knowledge on cell line- [56] and tissue-specific [259] lysosomal protein expression levels significantly, and confirmed the notion on lysosomal (protein) heterogeneity as well as lysosome diversity [106, 257], which was already speculated decades ago [260-262]. A major drawback of current MS-based experiments, alongside their dependence on lysosome enrichment which limits investigation to selected cell lines and tissue types [9], is their incapacity to investigate the stoichiometric composition of the lysosomal proteome. Considering the various functional roles of different cell types and the organ-specific phenotypes of LSDs [35], reinforces the idea of a cell or tissue type-specific lysosomal proteome composition that is adapted to the specific functional role of the lysosome within the cell/tissue-type [56]. Since, until now, no absolutely quantified peptide standard has been available, determining the absolute quantity of lysosomal proteins and their stoichiometric relationships in different cells and tissues has not been possible. MS-based proteomics of lysosomal proteins has been confined to relative comparisons between the same or similar sample types, minimizing sample matrix effects [171, 172] and thus limiting (relative) lysosomal proteomics comparisons to same tissues [259, 263, 264] and cell culture-based systems [56, 117, 241]. Knowledge of absolute lysosomal protein quantities would allow the mapping of quantitative similarities and differences in lysosomal proteome composition across various cell/tissue types. Protein stoichiometries of either the complete lysosomal proteome or of known lysosomal protein interaction partners and complexes could be studied. Furthermore, absolute protein quantities would facilitate the quantitative investigation of lysosomal protein distribution within a cell. Also, to enhance the understanding of LSDs and their underlying pathophysiology, stoichiometries of proteins involved in affected metabolic steps and pathways are of interest. Dynamics in these stoichiometries can provide immediate insights on the molecular level of a disease, for example upon disease progression or treatment intervention. Especially with regard to the study of human LSDs,

established animal models play an important role, and are preferred over cell culture-based models due to their ability to reflect the pathophysiological phenotype of human LSDs more realistically [48, 265]. However, despite the numerous advantages and availability of animal models, the majority of tissues have remained inaccessible to a systematic, large-scale investigation of the lysosomal proteome using current (untargeted) MS-based proteomics experiments due to high sample complexity. In Chapter 2, it was previously demonstrated that applying a targeted MS approach, such as PRM enables the quantification of the lysosomal proteome even from highly complex sample types, thereby enabling a systematic study from non-enriched sample material.

In this final chapter, the results of the absolute quantification of lysosomal proteins from one immortalized cell line, four primary cell types, and one lysosome-enrichment experiment are presented. Primary cell types served as well-suited sample types for quantitative characterization as well as for benchmarking the developed targeted MS platform and its integration with a stable isotope-labeled, absolutely quantified internal peptide standard. Unlike immortalized cell lines, primary cells retain the initial characteristics and cellular functions of their tissue of origin, making them ideal for studying cell/lysosomal biology and disease mechanisms [266-268]. However, they yield comparatively lower protein amounts compared to immortalized cell lines, which typically excludes them from enrichment protocols. This limitation makes a targeted, enrichment-free MS analysis strategy an ideal choice for the analysis of primary cell type samples. For absolute quantification a stable isotope-labeled, absolutely quantified internal peptide standard was generated applying the QconCAT strategy [206]. The engineered artificial standard proteins (QconCATs), harboring the internal peptide standards, represent 144 lysosomal proteins, which comprised the majority of the lysosomal proteome at the start of the project in 2016 [58]. To achieve an enrichment-independent and unbiased absolute quantification of lysosomal proteins from whole cell lysates (WCL), the MRM approach [135], often referred to as the gold standard for quantification [269], was selected for targeted MS-based absolute quantification.

5.2 Material and methods

5.2.1 Peptide selection and construction of QconCAT protein standards

Based on analyses of in-house and published datasets [270-273] ≥ 2 peptides/lysosomal protein were manually selected from a pool of > 6300 unique tryptic peptides considering the following factors: a) proteotypicity, b) signal intensity, c) possibility of PTMs, d) missed cleavage sites, e) proximity to proline, acidic amino acids or protein termini, and f) cleavage between two basic residues. If more than three peptides per lysosomal protein matched the criteria, the three peptides with the highest signal intensity were chosen. Selected lysosomal protein-derived peptides were grouped based on their abundance (averaged across individual peptides) in large scale proteomics datasets of brain and liver [270, 271, 273] into 12 groups (from highest intensities in QconCAT 1 to lowest intensities in QconCAT 12) with all peptides representing a given lysosomal protein on the same QconCAT. Subsequently, the order of peptides in each QconCAT was optimized *in silico* via PeptideCutter [274] to achieve maximum cleavage efficiency. To enable absolute quantification of individual QconCATs, a MIPA peptide [201] was placed at the N-terminus. A 6x His-tag was placed at the C-terminus for Ni^{2+}affinity purification, of each QconCAT.

5.2.2 Plasmid design and expression of stable isotope-labeled QconCATs

Designed QconCAT protein sequences were reverse-translated, codon-optimized, generated by gene synthesis, and cloned into pET-28a (+) expression vectors (69864-3, EMD Biosciences). To enhance bacterial expression efficiency of lysosomal QconCAT 12, it was further fused with maltose-binding protein (MBP, [275]) and the MBP-QconCAT 12 fusion construct was sub-cloned into the PBAD-his6-prkA-pACYC184 (pRKA) expression vector [276]. pRKA was a gift from Ichiro Matsumura (Addgene plasmid # 41041; http://n2t.net/addgene:41041; RRID: Addgene_41041). Individual QconCAT plasmids were transformed into *E. coli* BL21 DE3 cells and stable isotope-labeled protein constructs were expressed and purified as described elsewhere [206]. Briefly, bacteria were grown in M9 minimal medium (including kanamycin for QconCATs 1-11 and ampicillin for QconCAT 12), containing $^{13}C_6{}^{15}N_4$ arginine (CNLM-539-H-0.05) and $^{13}C_6{}^{15}N_2$ lysine (CNLM-291-H-0.05, both Cambridge Isotope Laboratories) until an optical density (OD) of 0.6-0.8 was reached. Subsequently QconCAT protein expression was induced by addition of 1 mM (final concentration) isopropyl β-D-1-thiogalactopyranoside (IPTG, Thermo Fisher Scientific), and bacterial cultures were incubated without agitation at room temperature (RT) for 4-6 hrs until reaching a final OD of 0.8-1. Bacteria were harvested by centrifugation at 8000x g, 4 °C for 10 min, the supernatant was discarded, and cell pellets were stored at -20 °C. For cell lysis, bacteria were resuspended in 2.5 ml of B-PER bacterial lysis buffer (Thermo Fisher Scientific) per 1 g of cell pellet and centrifuged at 6000x g, 4 °C for 10 min. Supernatant was discarded, and the pellet was resuspended in 2.5 ml fresh

B-PER. Lysozyme (0.2 mg/ml, Sigma Aldrich), protease inhibitor cocktail (1x cOmplete EDTA-free, Roche) and DNaseI (300 U, Thermo Fisher) were added (all final concentrations) and the bacterial lysate was incubated on a horizontal shaker at RT for 10 min. Subsequently, 15 ml of a 1:10 B-PER dilution (in water) was added and the solution was vortexed and centrifuged at 15,000x g, 4 °C for 15 min. The pellet containing bacterial inclusion bodies was washed twice with 20 ml of a 1:10 B-PER dilution and stored at -20 °C until further use. Inclusion bodies were solubilized with either 1% sodium dodecyl sulfate (SDS) or 6 M guanidinium chloride (GCl), in a buffer containing 500 mM NaCl, and 20 mM sodium dihydrogen phosphate (NaH_2PO_4)/disodium hydrogen phosphate (Na_2HPO_4), adjusted to pH 7.4. Samples were diluted to a final volume of 20 ml with 0.1% SDS or 6 M GCl, centrifuged at 15,000x g, RT for 15 min and purified by pre-equilibrated nickel-charged 1 ml His-trap columns (17-5247-01, GE Healthcare). Bound proteins were eluted using 500 mM NaCl, 20 mM NaH_2PO_4/Na_2HPO_4, 500 mM imidazole, and 50 mM ethylenediaminetetraacetic acid (pH 7.4) in 0.1% SDS or 6 M GCl. Protein eluate was stored at -20 °C.

5.2.3 Cell culture and lysosome enrichment

All cell culture experiments were performed under sterile conditions and cells were cultured at 37 °C, 100% humidity, and 5% CO2. MEFs were seeded at a density of $1.5x10^6$ cells and cultured for 72 hrs using Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum (FBS), 100 mU/ml penicillin, 100 µg/ml streptomycin, and 2 mM glutamine (all Thermo Fisher Scientific). For macrophage (MP), osteoclast (OC) and osteoblast (OB) differentiation, the bone marrow was flushed out of the femora from six wild-type mice (C57BL/6 mouse) per cell type at the age of 12 weeks with α-minimal essential medium (α-MEM) containing 10% FBS. The care and use of mice complied with all relevant local animal welfare laws, guidelines and policies (ORG-1091). Animal experiments were approved by the animal facility of the University Medical Center Hamburg-Eppendorf and by the Behörde für Gesundheit und Verbraucherschutz. Wild-type C57BL/6 mice were housed in a pathogen-free facility at the University Medical Center Hamburg-Eppendorf, with a 12-hour light/dark cycle, 45% to 65% relative humidity, and an ambient temperature of 20 °C to 24 °C. Experimental procedures were conducted following institutional guidelines. Extracted cells were plated at a density of $12.5x10^6$ cells per well, incubated for 24 hrs, the media was removed, and dead cells were washed off. Adherent cells were further cultured in α-MEM containing 10% FBS and 10 nM 1,25-dihydroxyvitamin-D3 (Sigma Aldrich). Beginning at day 4 after seeding, M-Csf (recomb. murine, 315-02, Peprotech) and receptor activator of NF-κB ligand (recomb. murine *E. coli* derived, 315-11, Peprotech) were added to a final concentration of 20 ng/ml and 40 ng/ml, respectively, and cells were cultured for 7 days to generate osteoclasts. Primary macrophages were generated by the same method without addition of Rankl (NF-κB ligand). For osteoblast differentiation, bone marrow cells were cultured for 10 days in α-MEM containing 10% fetal bovine serum, 25 µg/ml L-ascorbic acid

(Sigma Aldrich), and 5 mM ß-glycerophosphate (Sigma Aldrich). For lung fibroblasts (LFs), lungs of six P12 wild-type mice were dissected and enzymatically digested in 1x phosphate-buffered saline (PBS) containing collagenase type III (150 U/ml, PAN Biotech) and dispase I (2.5 U/ml, Sigma-Aldrich). After incubation at 180 rpm, 37 °C for 3 hrs, strained cells were seeded ($12.5x10^6$ cells per well) and cultured for 7 days. For harvesting, cells were washed with ice-cold 1x PBS, scraped in 500 µl ice-cold PBS, and centrifuged at 1000x g, 4 °C for 10 min. Supernatants were discarded and the cell pellet was stored at -80 °C. Lysosomes were enriched from MEFs using SPIONs from two 10 cm plates with a seeding density of $3x10^6$ cells as described elsewhere [277]. Briefly, cells were incubated with SPIONs for 24 hrs, followed by 24 hrs chase, harvested by scraping, lysed using a dounce homogenizer, and lysosomes were enriched from the postnuclear supernatant (PNS) using LS-columns and a magnetic stand (Miltenyi Biotec). Lysosome-enriched fractions (LEFs) were concentrated by centrifugation at 20,000x g, 4° C for 30 min and the supernatant was discarded. For the determination of cellular protein content in MEFs, cells were harvested by trypsinization, cell numbers were determined using a Neubauer chamber, and the protein content of defined amounts of cells was determined using the DC protein assay (Bio-Rad).

5.2.4 Proteolytic digestion of QconCATs

To evaluate the performance of different digestion strategies on peptide release from QconCATs, proteins eluted in 0.1% SDS were digested by SP3, RapiGest, or IGD, and such eluted in 6 M GCl in solution using GCl or FASP, as described elsewhere [224, 226, 232, 278, 279] with minor modifications. In the following, all approaches are briefly outlined. For reduction/alkylation reagents final concentrations are indicated. For SP3, 10 µg of protein were adjusted to a total volume of 10 µl with 1% SDS, 50 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES, pH 8) in PCR-tubes. Proteins were reduced with 5 mM dithiothreitol (DTT) at 56 °C for 45 min, alkylated with 20 mM acrylamide at RT for 30 min, and the reaction was quenched with 5 mM DTT at RT for 10 min. SP3 beads (1:1 mixture of Sera-Mag SpeedBeads 45152105050250 and 65152105050250, GE Healthcare) were added at a 10:1 (w/w) bead to protein ratio and protein binding was induced by addition of ethanol (final concentration 80%) and incubation at 1000 rpm, RT for 20 min on a thermomixer. Using a magnetic rack, the supernatant was discarded, the particle-bound proteins were washed (2x with 150 µl 80% ethanol and 1x with 100 µl 100% acetonitrile (ACN)), and the supernatant removed. Subsequently, proteins were proteolytically digested by addition of 50 µl 100 mM ammonium bicarbonate solution containing trypsin at a 1:50 (w/w) protease-to-protein ratio and incubation in a thermomixer at 1000 rpm, 37 °C overnight. The next day, beads were immobilized using the magnetic rack and the peptide-containing supernatants transferred to fresh tubes. For RapiGest digestion, 0.25 µg of protein were resuspended in 1% RapiGest (Waters), diluted 1:4 with 0.133 M ammonium bicarbonate solution to a final volume of 20 µl, reduced with 5 mM DTT at 56 °C

for 45 min, alkylated with 20 mM acrylamide at RT for 30 min, and the reaction was quenched with 5 mM DTT for 10 min at RT [231, 232]. Protein digestion was performed at 37 °C overnight using trypsin at a protease-to-protein ratio of 1:5 (w/w) in a final volume of 30 µl, and a final concentration of 0.1% RapiGest in 100 mM ammonium bicarbonate. The next day, RapiGest was hydrolyzed by addition of 1% (final concentration) trifluoroacetic acid (TFA) and incubation at 800 rpm, 37 °C for 45 min. RapiGest hydrolysis products were precipitated by centrifugation at 14,000x g, RT for 15 min and peptide-containing supernatants were transferred to fresh tubes. For in gel digestion, 75 µg of protein were combined with 4x Laemmli loading buffer (8% (w/v) SDS, 10% (v/v) β-mercaptoethanol (BME), 40% (v/v) glycerol, 0.04% (w/v) bromophenol blue, 0.5 M tris(hydroxymethyl)aminomethane-hydrochloric acid (Tris-HCl), pH 6.8) to reach a final concentration of 1x [280]. Samples were reduced at 95 °C for 10 min, alkylated with 20 mM acrylamide at RT for 30 min, and the reaction was quenched with 5 mM DTT at RT for 10 min. A sample volume corresponding to 60 µg protein was distributed across 9 lanes of a 10% polyacrylamide gel (~6.7 µg/lane) and SDS-PAGE was performed at a constant voltage of 120 V for 90 min. Gels were washed with milliQ water, and stained overnight using Coomassie brilliant blue (Thermo Fisher Scientific). The next day, gels were washed three times with milliQ water, the QconCAT bands excised and cut to small pieces, pooled, and transferred to a 1.5 ml tube. Gel pieces were de-stained using 30% ACN 70 mM ammonium bicarbonate solution, dehydrated with ACN, and dried using a vacuum centrifuge. Subsequently, gel pieces were rehydrated with 100 mM ammonium bicarbonate containing trypsin at a protease-to-protein ratio of 1:25 (w/w), digested overnight at 37 °C and the peptides recovered the next day by successive incubation of the gel pieces with 50% ACN 0.1% TFA, 100 mM ammonium bicarbonate, and 50% ACN 50 mM ammonium bicarbonate. Recovered peptide-containing supernatants were pooled and dried using a vacuum centrifuge. For FASP, all incubation steps were followed by centrifugation of the filter unit at 14,000x g, RT for 15 min and discarding of the flow through unless stated otherwise. For each sample, 20 µg of protein was brought to a final volume of 300 µl with 6 M GCl, 100 mM Tris-HCl (pH 7.5). Samples were loaded on equilibrated centrifugal filter units (Microcon-30kDa, MRCF0R030, Merck) and samples were concentrated at 14,000x g, RT for 25 min. Subsequently, the filter units were washed twice with 100 µl of 6 M GCl, 100 mM Tris-HCl. Proteins were reduced with 10 mM BME at 56 °C for 45 min, and alkylated with 10 mM chloroacetamide at RT for 30 min. Centrifugal filter units were washed three times with 100 µl Tris-HCl (pH 7.5) and twice with 100 µl 100 mM ammonium bicarbonate solution (pH 8). Samples were digested by addition of trypsin in 50 µl 100 mM ammonium bicarbonate at a protease-to-protein ratio of 1:100 (w/w) and incubation at 37 °C overnight in a wet chamber. Peptides were recovered by centrifugation and such retained on the filter membrane eluted by an additional centrifugation step with 50 µl 500 mM NaCl. For GCl in solution digestion, an equimolar mix of QconCAT proteins (1 µg per QconCAT protein) was prepared and adjusted to 30 µl of 5.5 M GCl in 100 mM HEPES (pH 7.6). Proteins were reduced with 10 mM BME at 56 °C for 45 min, alkylated with 20 mM chloroacetamide at RT for 30 min, and the reaction

was quenched with 10 mM BME at RT for 10 min. For proteolytic digestion the GCl concentration was further reduced to 1 M, trypsin was added at a 1:5 (w/w) protease-to-protein ratio, and samples were incubated overnight at 37 °C.

5.2.5 Proteolytic digestion of cell pellets and lysosome-enriched fractions

Cell/lysosome pellets were resuspended in 4% SDS, 0.1 M HEPES (pH 8.0) at a pellet to buffer ratio of 1:10 (v/v), heated at 95 °C for 10 min, and sonicated three times for 30 s at 60% duty cycle and an output of 6 with ultrasonic processor (MSK-USP-3, 300W) on ice until complete solubilization of pellet was observed. Lysates were incubated again at 95 °C for 5 min, centrifuged at 20,000x g, RT for 30 min, and the clear supernatants were transferred to fresh tubes followed by determination of protein concentrations using the DC protein assay. Samples were stored at -80 °C until further use. For each sample, 100/250 µg of protein was precipitated with methanol/chloroform as described elsewhere [281], and pellets were solubilized in 0.2% RapiGest, 200 mM ammonium bicarbonate (pH 8) by incubation at 400 rpm, 95 °C for 15 min, ultrasonication for 30 min at RT in a water bath, and pre-digestion with trypsin at 37 °C for 45 min at a protease-to-protein ratio of 1:500. Subsequently samples were diluted with water (0.15% RapiGest, 150 mM ammonium bicarbonate (pH 8)), proteins were reduced using 5 mM DTT at 56 °C for 45 min, alkylated with 20 mM acrylamide at RT for 30 min, and the reaction was quenched with 5mM DTT (all final concentrations) [282]. Samples were adjusted to 0.1% RapiGest at protein concentration of ≥ 1µg/µl [231] and trypsin was added at a protease-to-protein ratio of 1:50. Proteolytic digestion was performed in a thermomixer at 500 rpm, 37 °C overnight. Samples were acidified with 1% TFA (final concentration), incubated at 37 °C for 45 min, centrifuged at 14,000x g, RT for 15 min, and the supernatants were transferred to a fresh tube.

5.2.6 Peptide desalting and quantification

Peptides from digests with inputs of > 75 µg were desalted with OASIS HLB cartridges (1cc 10mg, Waters) and such with < 75 µg with C_{18} STAGE tips (Empore discs, # 2215, 3M) as described elsewhere [283]. Briefly, tips were activated with methanol and cartridges/tips washed with and 80% ACN, 0.5% acetic acid (AA) and 0.5% AA. Samples were acidified with 0.5% AA (final concentration), loaded to the cartridge/tip, washed with 0.5% AA, and eluted with 80% ACN, 0.5% AA. Eluate fractions were dried using a vacuum centrifuge, resuspended in 5% ACN, and peptide amounts were determined using a fluorometric peptide assay (23290, Thermo Fisher Scientific).

5.2.7 Liquid chromatography tandem mass spectrometry data acquisition

The following liquid chromatography tandem mass spectrometry setups were used: an EASY-nLC 1000 UHPLC coupled to an LTQ Orbitrap Velos (both Thermo Fisher Scientific), a Dionex Ultimate 3000 UHPLC coupled to an Orbitrap Fusion Lumos (both Thermo Fisher Scientific), a nano ACQUITY UPLC system (Waters) coupled to a QTRAP 5500 (Sciex) and a Dionex Ultimate 3000 UHPLC coupled to a QTrap 6500+ (Sciex). If not indicated otherwise, spray tips were produced in-house from fused silica capillaries (100 µm inner/360 µm outer diameter) with a laser puller (P-2000, Sutter Instruments) and packed with either 5 µm or 3 µm C_{18} particles (Reprosil-Pur 120 C_{18} AQ, both Dr. Maisch).

For LTQ Orbitrap Velos analyses, peptides were loaded onto the analytical column (20 cm length, 5 µm particles) with 99% buffer A (5% dimethyl sulfoxide (DMSO), 95 % H_2O, 0.1% formic acid (FA)) 1% buffer B (95% ACN, 5% DMSO and 0.1% FA) at 1 µl/min for 20 min and eluted with a 30-min linear gradient from 1-35% buffer B at a flow rate of 400 nl/min. Eluting peptides were ionized in the positive ion mode and MS1 scans were recorded in the OT analyzer from m/z 300-2000 at a resolution of 60,000. The ten most abundant MS1 precursor ions were selected for CID fragmentation with a normalized collision energy (NCE) of 35, MS2 spectra were acquired in the linear ion trap, and dynamic exclusion was set to 30 s. For Orbitrap Fusion Lumos analyses, samples were either loaded directly to the analytical column (35 cm, 3 µm particle size) with 99% buffer A (100% H_2O, 0.1% FA) 1% buffer B (90% ACN, 10% H_2O, 0.1% FA) at a flow rate of 600 nl/min for 25 min, or onto a trapping column (Acclaim PepMap C_{18} 5 µm 100 Å, 160454, Thermo Fisher Scientific) with 0.5% TFA at a flow rate of 10 µl/min for 8 min. Peptides were eluted with linear gradients of 30-, 60- or 120 min from 1-35% buffer B at 300 nl/min and ionized in the positive ion mode. For DDA analyses, MS1 scans were acquired from m/z 250-2000 at a resolution of 60,000 in the OT analyzer with an automatic gain control (AGC) target setting of $4x10^5$ and a maximum injection time set to 50 ms or to “automatic”. The most abundant MS1 precursor ions (+2 to +7) were isolated in the quadrupole (m/z 1.2-1.6), fragmented by higher energy collisional dissociation with an NCE of 30 in the top speed mode (2-3 s cycle time), and excluded for 15-60 s from fragmentation. MS2 fragment ion spectra were acquired at a resolution of either 15,000 or 30,000 in the OT mass analyzer using a AGC target setting of $5x10^4$ with maximum injection times between 22-100 ms. For PRM experiments, MS1 scans were recorded from m/z 200–2000 and the maximum injection time set to 118 ms. Up to 140 ms maximum injection time was used in the PRM experiment, recording MS2 spectra at a resolution of 60,000.

For MRM experiments conducted with the QTrap5500 mass spectrometer, peptides were injected at a flow rate of 10 µl/min in 99.4% buffer A (98.9% H_2O, 1% ACN, 0.1% FA) 0.6% buffer B (89.9% ACN, 10% H_2O, 0.1% FA) for 7 min onto a trapping column (Symmetry C_{18} 5 µm 100 Å, 186007496, Waters), followed by their elution onto, and separation by, an analytical column (M-Class Peptide

BEH C_{18} 1.7 µm,130 Å, 186007484, Waters). Peptides were eluted with a linear gradient of 3-40% buffer B for 177 min at a flow rate of 200 nl/min and a column temperature of 35 °C. For the QTrap 6500+ mass spectrometer, samples were loaded at a flow rate of 20 µl/min with 0.1% TFA for 4 min onto a trapping column (Acclaim PepMap C_{18} 5 µm 100 Å, 160454, Thermo Fisher Scientific) followed by their elution to, and separation by, the analytical column (35 cm, 3 µm particle size) at 300 nl/min with a linear gradient from 7% buffer B (90% ACN, 10% H_2O, 0.1% FA) 93% buffer A (100% H_2O, 0.1% FA) to 35% buffer B in 30 or 60 min. Both QTrap instruments were operated in the MRM mode (low mass region) using unit resolution for the quadrupole analyzers. Peptides were ionized in the positive mode. MIDAS experiments were performed in the unscheduled mode, with 20 ms dwell time per transition and a maximum of 60-75 transitions per sample injection [149]. CEs were optimized for a selected set of peptides (individual transitions, step size 2.5 V) using Skyline (21.1.0.146, skyline.ms [151, 284]), the default CE parameters and optimized values were incorporated to the assay [150]. Scheduled MRM experiments were conducted with 3 s cycle time and detection windows of 3 min defined.

5.2.8 MS data processing and analysis

Acquired DDA files were processed with Proteome Discoverer (2.5.0.400, Thermo Fisher Scientific) utilizing Mascot as search engine (v2.6.1, Matrix Sciences) or Mascot as standalone application. Searches were performed against a custom QconCAT database (Table S1), or a combination of the QconCAT database and SwissProt *E. coli* (4403 entries, release 07/2023), along with a database containing common contaminants [285], with a concatenated decoy generation strategy. For database searches, fully tryptic peptides with up to one (labeling efficiency study and library generation) or three (missed cleavage study) missed cleavage sites were considered. The mass tolerance was set to 10 ppm (MS1) and 20 ppm (MS2) for data acquired on the Orbitrap Fusion Lumos and 10 ppm (MS1) and 0.6 Da (MS2) for data acquired on the LTQ Orbitrap Velos. MRM data were converted to mgf files using ProteoWizard (3.0.221466 [286]) and searches performed with a 1.2 Da MS1 and 0.6 Da MS2 mass tolerance. For all searches, propionamide (C) or carbamidomethyl (C) were defined as static modifications and oxidation (M) as well as acetyl (protein N-term) as variable modifications. Heavy amino acid labeling with $^{13}C_6{}^{15}N_2$ (K) and $^{13}C_6{}^{15}N_4$ (R) was set as fixed modification for the analysis of biological samples and as variable modification for labeling efficiency analyses. Identified peptide PSMs were validated based on q-values with Percolator [287] using a 1%/5% (strict/relaxed) target false discovery rate (FDR). Combined PSMs were aggregated to peptides and proteins applying the principle of strict parsimony, and a 1% FDR filter was applied at the peptide and protein level. The stable isotope labeling efficiency and missed cleavage rate was determined for each protein construct individually at the peptide level, contaminant peptides were excluded for the analysis.

The stable isotope labeling efficiency rate was calculated as: 1-(1/(Heavy/Light)), whereas Heavy/Light is the abundance ratio of the heavy (stable isotope-labeled), over the light (non-stable isotope-labeled) precursor area under the curve (AUC) as reported by Proteome Discoverer. The average of all peptides per QconCAT protein construct was reported. The missed cleavage rate was calculated as: $\Sigma AUC_{MC}/(\Sigma(AUC_{MC}+AUC_{NMC}))$, reporting the protein construct's cumulative missed cleavage rate. AUC is the area under the curve of each peptide precursor in either its cleaved (NMC i.e. non-missed-cleaved) or miscleaved (MC) form.

MRM data were analyzed with Skyline. Spectral libraries were built with the BiblioSpec plugin of Skyline applying a cut-off score of 0.95 (settings: filter for documented peptides, exclude ambiguous matches). When both QTrap and Orbitrap MS2 spectra were available for one peptide, QTrap data were used. For peptides with no available spectra, MS2 spectra were generated *in silico* using PROSIT [68] (NCE 27, default settings with intensity prediction model: publication_v1). MRM data quality was monitored by manual inspection and by using the AutoQC loader application (22.1.0.122, [288]) in combination with Skyline. Sample-specific peak picking models were generated with mProphet [289] and trained using the 2nd best peak option, excluding negatively contributing features. Inconclusive stable isotope-labeled peptide standard traces were deactivated and excluded from the following data analysis. In the case of interfering or low/noisy endogenous sample peptide traces, the theoretically highest fragment ion trace was used, excluding other transitions („Not quantitative" feature). Skyline peptide quantity exports were processed in MS Excel and R (v4.2.2 [290]) using the reshape2 (v1.4.4), dplyr (v1.1.3), tidyr (v1.3.0), and stringr (v1.5.0) packages. For peptides with both oxidized and non-oxidized variants, the summed values of both versions were used. Signal-to-noise (S/N) ratios were calculated as ratio of total fragment/total background areas. Peptide quantifications were classified into three categories as follows: "A" for S/N ratios ≥ 3 of both endogenous and stable isotope-labeled peptide, "B" for S/N ≥ 3 of stable isotope-labeled peptide, and S/N < 3 for endogenous peptide, and "C" if stable isotope-labeled peptide had S/N < 3. C peptides were not considered for protein quantification, employing two modes for removal: medium and stringent. Medium removal excluded class C peptide quantifications in the affected sample/condition, requiring at least two remaining quantifications per sample/condition. Stringent class C peptide removal eliminates the peptide quantification entirely from the data matrix across all samples/conditions. Unless otherwise stated, median absolute protein amounts are reported for all replicates per sample and/or condition and the variance was assessed via the robust standard deviation (rSD), calculated as: 1.4826 x median absolute deviation (MAD) which is defined as: $(|X_i - \text{median}_j(X_j)|)$. The robust coefficient of variation (rCV) as relative measure of variance was calculated as: rSD/median [219]. For visualization the R packages ggplot2 (v3.4.3), gridExtra (v2.3), RColorBrewer (v1.1.3) and ComplexHeatmap (v3.2), as well as GraphPad Prism (10.1.2) were used.

5.2.9 Immunofluorescence analysis

The following experiment was performed by Sofía Fajardo-Callejon (Institute for Biochemistry and Molecular Biology, Rheinische Friedrich-Wilhelms-University of Bonn). Cells were seeded on 15 mm coverslips in 12-well plates at a density of $1.5x10^4$ and grown for 24 hrs. For Ctsd visualization, cells were incubated with SiRLysosome (1 µM, SC02, Tebu-bio) for 2 hrs prior to fixation. Cells were fixed in 4% paraformaldehyde (Sigma Aldrich), 1x PBS for 25 min at 37 °C, washed 3x with 1x PBS, permeabilized with 0.1% Saponin (S7900, Sigma Aldrich/Merck) 2% bovine serum albumin (BSA) in 1x PBS at RT for 10 min, and subsequently incubated with anti Lamp2 (1:400, clone ABL-93, DSHB) in blocking solution (0.1% Saponin, 2% BSA in 1x PBS) at 4 °C overnight. The following day, cells were washed (1x PBS) and incubated in blocking solution with fluorophore-conjugated secondary antibody (1:500, Cy3 AffiniPure Goat Anti-Rat, 112-166-062, Dianova) at RT for 1 hr. Washed cells were mounted with 1x ROTI Mount FluorCareDAPI (HP20.1, Carl Roth) to specimen slides. Fluorescence microscopy imaging was performed with an Axio Observer 7 microscope in combination with a 63x objective (Axiocam 705 mono, both ZEISS). Images were deconvoluted, processed with maximum intensity projection (v3.4, Zen Bluepro, ZEISS), and lysosomes were counted with segmentation masks generated using Ilastik [291]. A subset of images was manually annotated based on Lamp2 signals to train a custom classification model for lysosomes. The classifier was applied to the full dataset, and the Fiji package [292] was used for the definition of lysosome masks and lysosome counting. For mask definition, particle edge smoothing, hollow particle filling, and watershed segmentation was applied. The Mander's coefficient was calculated using the Fiji JaCoP plugin.

5.2.10 Quantitative RNA analysis of primary cell types

The following experiment was performed by Maryam Omidi and Sandra Pohl (Department of Osteology and Biomechanics, University Medical Centre Hamburg-Eppendorf). RNA was isolated from cultured lung fibroblasts, macrophages, and osteoblasts (Monarch Total RNA Miniprep Kit, NEB) and 1 µg of total RNA from pooled samples was reverse transcribed (High-Capacity cDNA Reverse Transcription Kit, Thermo Fisher Scientific). For quantitative PCR (qPCR), TaqMan gene expression assays with pre-designed probes and primers for lysosomal genes were used [293, 294]. All reactions were run in technical triplicates using the primaQUANT qPCR Master Mix (Steinbrenner) on a QuantStudio 1 PCR system (Thermo Fisher Scientific). Relative mRNA levels were normalized to Gapdh within the same cDNA sample, using the $2^{-\Delta CT}$ method. ΔCT is defined as: CT(lysosomal gene) – CT (Gapdh)

5.2.11 Absolute quantification and protein copy number determination

Absolute quantification of the generated QconCAT standards through liquid chromatography multiple reaction monitoring mass spectrometry (LC-MRM-MS) was performed with a minimum of three injection replicates, with replicate values averaged. Acquired data was manually evaluated in Skyline, adjusting integration borders of the QconCAT construct-specific, stable isotope-labeled MIPA signals. Light/Heavy ratios for each individual stable isotope-labeled MIPA were derived with respect to its non-stable isotope-labeled MIPA counterpart using the exported total fragment area values. The unknown quantities of the stable isotope-labeled MIPA, hereafter referred to as amount (Heavy), respectively the associated lysosomal peptide standards per stable isotope-labeled QconCAT protein construct were calculated as followed:

A)

$$\text{amount (Heavy)} = \frac{\text{amount (Light)}}{\text{ratio AUC (Light/Heavy)}}$$

with: amount (Light) as the known non-stable isotope-labeled peptide amount injected (in fmol) for the experiment, and ratio AUC (Light/Heavy) as the ratio of the total fragment area between the non-stable isotope-labeled (Light) and the stable isotope-labeled (Heavy) MIPA variant derived from the MS-experiment.

Endogenous amounts of lysosomal peptides were calculated as followed:

B)

$$\text{amount (Light)} = \text{amount(Heavy) x ratio AUC (Light/Heavy)}$$

with: amount (Heavy) as the known stable isotope-labeled peptide amount injected (in fmol) for the experiment, and ratio AUC (Light/Heavy) as the ratio of the total fragment area between the non-stable isotope-labeled endogenous peptide (Light) and the stable isotope-labeled lysosomal peptide standard (Heavy) derived from the MS-experiment.

Protein copy numbers of individual proteins were determined using the following equation:

C)

$$\text{protein copy number} = \text{protein amount x N}$$

whereas N is defined as the Avogadro constant (6.02214076 x 10^{23} mol^{-1}), and the protein amount as experimentally determined (in fmol).

For the determination of individual protein copy numbers from MEF whole cell lysate (WCL) per one cell, the equation in C) was modified to:

D)

$$\text{protein copy number per cell (WCL)} = \frac{\text{protein amount x N}}{\text{number cells}}$$

whereas the number cells refers to the cell count (65,396 cells) which was obtained by combining information on the injected protein amount for the MS experiment, as well as the protein content per single cell (0.22 ng/cell, Figure S4).

Individual protein copy numbers in lysosome-enriched fractions (LEF) per one lysosome were calculated as:

E)

$$\text{protein copy number per lysosome (LEF)} = \frac{\text{protein copy number (LEF)}}{\text{number lysosomes (LEF)}}$$

whereas the number of lysosomes in the LEF, here referred as number lysosomes (LEF), was determined as:

F)

$$\text{number lysosomes (LEF)} = \frac{\text{total protein copy number Lamp2 (LEF)}}{\text{Lamp2 protein copy number per lysosomes (WCL)}}$$

with: Lamp2 protein copy number per lysosomes (WCL) being determined as 1400 (based on median mouse embryonic fibroblast WCL data) and an average of 482 lysosomes per cell (Figure 6B).

5.2.12 Protein distribution analysis

For the correlation and comparison of protein copy numbers obtained from the two individual experiments (LEF and MEF), lysosomal protein copy numbers from WCL data had to be normalized per one lysosome.

Normalized protein copy numbers for MEF WCL samples were calculated as:

G)

$$\text{protein copy number per lysosome (WCL)} = \frac{\text{protein copy number (WCL)}}{(\text{total number cells x number lysosomes per cell})}$$

with: total number cells (65,396 cells) determined from cell counting and protein assays, and number lysosomes per cell as determined from immunofluorescence analysis (482 lysosomes per MEF cell, Figure 6B).

The number of non-lysosomal located protein copies per normalized unit (one lysosome) was defined as:

H)

$$\text{protein copy number per lysosome (WCL)} - \text{protein copy number per lysosome (LEF)}$$

Consequently, for proteins localizing (almost) exclusively in the lysosome, normalized protein copy numbers should be similar in both LEF and WCL samples. Proteins with additional other subcellular, i.e. non-lysosomal localization, should present with larger protein copy number values in WCLs – which is a combination of lysosomal and non-lysosomal protein content – as compared to the LEF. In cases where the calculated protein copy number per lysosome (LEF) exceeded the protein copy number per lysosome (WCL), a complete lysosomal localization was assumed adjusting the protein copy number for the WCL experiment to the protein copy number determined in the LEF experiment. The relative proportional distribution (%) for lysosomal-located (as determined in E) and/or non-lysosomal located protein copy numbers (as determined in H) was calculated as proportion of the (total) protein copy number per lysosome (WCL) determined in G.

5.3 Result and discussion

In order to enable the absolute quantification of the lysosomal proteome across different types of mouse samples, a strategy based on the generation of stable isotope-labeled standard peptides and a targeted MRM-MS assay was developed. Proteins selected for this study had previously been confirmed to localize in either the lysosomal lumen or membrane, respectively.

5.3.1 Generation of absolute quantified protein standards

Based on a manually curated list of known lysosomal proteins [58] and proteins which were identified in screenings to be located at or in lysosomes (Table S1), 155 proteins were chosen. For those, tryptic surrogate peptides from both in house and published datasets [270-273] were selected, identifying > 6300 peptide candidates. Based on their biochemical and physiochemical criteria relevant for standard peptide release from QconCATs and MS detection, 414 peptides covering 144 proteins were manually selected (Figure 1A, Table S1, for selection criteria see Material and methods 5.2.1). Subsequently, peptides were grouped based on their average abundance in large scale proteomics studies [270-273] and distributed to 12 QconCAT proteins. Each QconCAT fusion protein contained peptides of lysosomal proteins with similar abundance. Then, the peptide order was optimized to yield full tryptic digestion of individual QconCATs. In addition, a C-terminal MIPA [201] as an internal standard for absolute quantification (Table S1), as well as an N-terminal His-tag for affinity purification, was fused to the constructs. Following the reverse translation of the final amino acid sequence and optimization for bacterial expression, synthetic constructs were generated by gene synthesis and cloned into bacterial expression vectors (Supplementary file). Finally, individual QconCATs were expressed in minimal media supplemented with stable isotope-labeled amino acids, followed by purification and quality control of the final protein expression using SDS-PAGE and western blotting (Figure S1).

The application of QconCAT-derived standard peptides for absolute quantification experiments requires the complete and efficient release of the individual reference peptides contained in the QconCAT fusion protein. Therefore, the performance of different digestion strategies including bead-based SP3 [226], in solution digestion using GCl [279] or RapiGest [231, 232], FASP [278], and IGD [224] was evaluated. Based on untargeted LC-MS/MS analyses of tryptic digests of all 12 QconCATs, the average missed cleavage rate was determined, identifying lowest values with SP3, RapiGest, and IGD (Figure 1B, Figure S2). The average labeling efficiency of fusion proteins was > 98% (Table S1). As IGD enables further to select QconCATs based on the molecular weight of the protein, eliminating truncated protein standards that result in false abundances due to the lack of distinct peptides, this strategy was chosen for the generation of stable isotope-labeled lysosomal peptide standards (SILLPS). Due to its compatibility with large sample amounts, RapiGest digestion

was selected for the proteolytic digestion of biological samples. To determine the absolute quantity of individual QconCAT-derived SILLPS, absolutely quantified synthetic MIPA reference peptides were obtained and used as internal standards for quantification in MRM analyses (Figure 1C, Table S1, Figure S3).

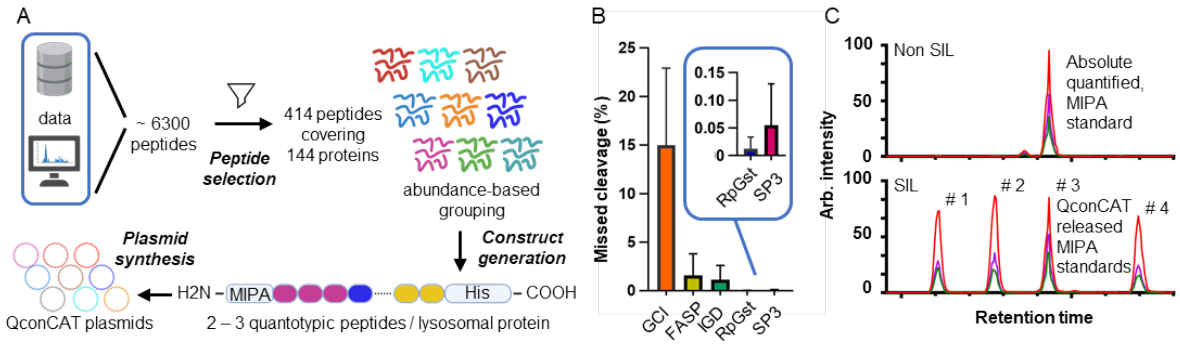


Figure 1: Design, evaluation and quantification of QconCAT protein standards. A) Schematic overview of the QconCAT workflow designing artificial QconCAT fusion proteins. Key steps include the selection and assembly of proteotypic peptides into a protein construct optimized for efficient proteolytic digestion. Following construct optimization and modification, synthetic genes are generated. B) Comparison of strategy-specific digestion performance. Reported are the cumulative missed cleavage rates (%) averaged for all 12 QconCAT protein standards per digestion strategy tested. C) Schematic representation of a multiplexed, absolute quantification experiment allowing quantification of four digested stable isotope-labeled (SIL) QconCAT proteins with one non SIL synthetic minimally permutated peptide analog (MIPA) peptide. Abbrev.: MIPA: minimally permutated peptide analog, SIL: stable isotope-labeled. Figure 1 A was generated using BioRender.

5.3.2 MRM assay development

For MS assay development master mixes containing equimolar amounts of all QconCAT-derived SILLPS were prepared, collision energies for peptide fragmentation were optimized, reference MS2 spectra for spectral library generation using DDA, PRM, and MIDAS [149] runs were defined, and the three most abundant, interference-free fragment ions were selected for the final MRM assay. Further, peptide-specific, normalized retention times using the iRT concept [148] were determined. To establish the SILLPS spike-in and on-column sample loading amounts, mixtures of cell line digests (fibroblasts, osteoblasts, osteoclasts, and macrophages) were utilized. For accurate quantification with the QTrap6500+ mass spectrometer, a range of up to two orders of magnitude between SIPPLS and endogenous peptides was aimed for. This resulted in two different spike-in doses of either 7.5 fmol or 22.5 fmol for the individual QconCAT-derived SILLPS (Figure 2A). Finally, optimal MS parameters, MRM assay size, gradient lengths, and numbers of runs were determined to cover the full assay. Peptides which did not meet the criteria for robust quantification in these analyses (see Material and methods) were removed from the assay.

The final MRM assay targets 388 lysosomal protein-derived peptides and their respective SILLPS covering > 95% of the included 143 lysosomal proteins with two or three surrogate peptides through ~2400 individual transitions. To enable acquisition of a sufficient number of data points across individual chromatographic peaks, the assay was split into four separate LC-MRM-MS analyses, with retention time windows as well as cycle and dwell times defined accordingly (Figure 2B, Table S1).

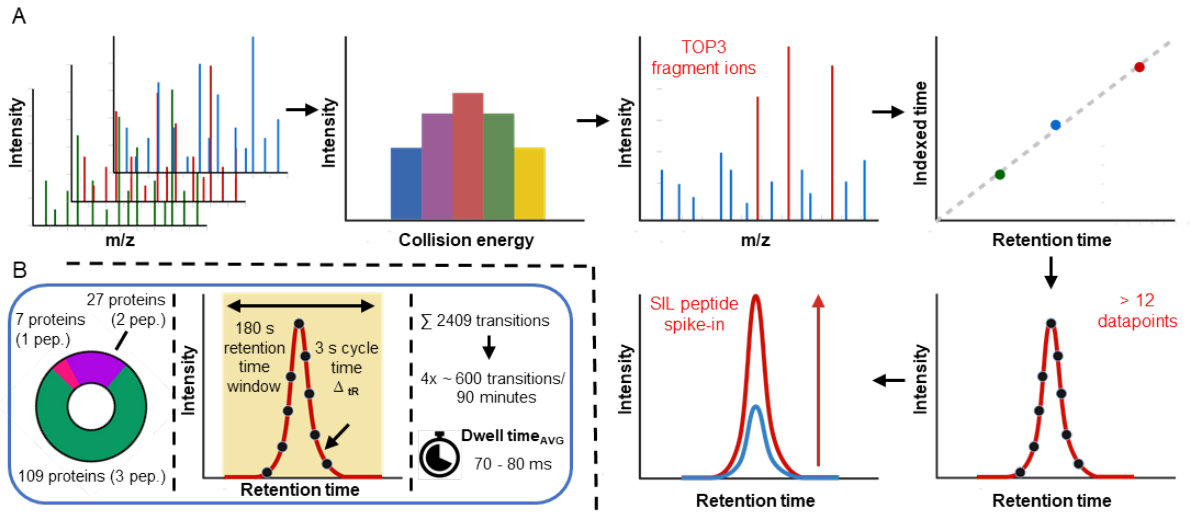


Figure 2: MRM assay development and optimization. A) Empirical data from MS experiments gathered are compiled into reference libraries (MS2 spectrum and normalized retention time). Targeted MS assay development comprises the determination and optimization of peptide, as well as assay-specific parameters in a multistep process. B) Key metrics of final MRM assay. Shown is the protein coverage by the number of peptides as well as assay-specific parameters used for data acquisition. Abbrev.: SIL: stable isotope-labeled. Figure 2 A was generated using BioRender.

5.3.3 Lysosomal protein copy number determination in Mouse Embryonic Fibroblasts

Many factors are crucial for correct lysosomal function. In order to ensure functional and stable organelles, several types of lysosomal proteins are needed. Structural proteins with heavily glycosylated luminal domains, such as Lamp1/2 [161] or Glmp [295], stabilize the lysosomal membrane [296] and prevent lysosomes from self-digestion [297]. The vacuolar ATPase (vATPase [298]), a 1 MDa multi protein complex, pumps protons across the lysosomal membrane, generating its acidic pH. For the export of degradation products to the cytosol and the bidirectional transport of ions such as Ca^{2+}, K^{+} or Cl^{-} a variety of membrane-embedded transporters, channels, and exchangers exist [299-301]. The degradative function of the lysosome, on the other hand, depends on a multitude of luminal hydrolases and their co-factors [302, 303], which can be broadly classified based on their substrates (e.g. proteases, lipases, nucleases, etc.) [304]. In order to determine the absolute quantity of individual lysosomal proteins covering multiple protein classes, the assay was initially used to determine lysosomal protein copy numbers in whole cell proteomes of MEFs, one of the most commonly used immortalized mouse cell lines (Figure 3).

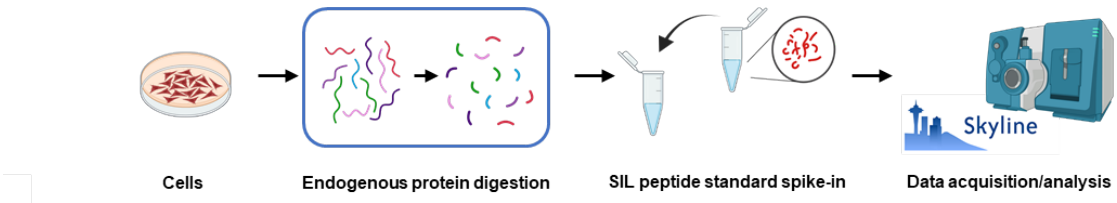


Figure 3: Workflow for absolute quantification experiments. Harvested cells are proteolytically digested and defined peptide amounts are mixed with known amounts of stable isotope-labeled (SIL) QconCAT peptide standard. Endogenous peptide/protein amounts are derived from absolute quantified peptide standard amounts. Abbrev.: SIL: stable isotope-labeled. The figure was generated using BioRender.

Protein copy numbers per cell were calculated by quantifying the average protein content of MEF cells, which was found to be 0.22 ng/cell (Figure S4), using a combination of cell counting and protein assays. This information was further combined with the SILLPS spike-in-derived values (Table S2, see Materials and methods). The resulting copy number/cell information for 141 MEF lysosomal proteins revealed a dynamic range of more than three orders of magnitude between the lowest and most abundant lysosomal proteins, respectively (Figure S5). With 2,000,000 copies per cell the Pituitary tumor-transforming gene 1 protein-interacting protein (Pttg1ip [305]) is the most abundant lysosomal protein in MEF. Among the highest abundant proteins are the lysosomal membrane proteins Lamp1 (840,000 copies/cell) and Lamp2 (675,000 copies/cell), as well as Npc2 - a small soluble protein essential for lysosomal cholesterol export - present at approximately 695,000 copies per cell (Figure 4 upper panel). Notably, these are the only proteins in the dataset that individually exceed 500,000 copies per cell, an observation that aligns closely with other studies identifying them as highly abundant [56]. Based on these copy number determinations, Lamp1 and Lamp2 constitute

together ~19% of lysosomal membrane protein molecules, equivalent to ~22% of lysosomal membrane protein biomass and 9% of total lysosomal biomass, respectively. This supports the role of these two highly glycosylated membrane proteins for the protection of the lysosomal membrane from attack by soluble hydrolases [161]. Among the soluble lysosomal hydrolases are Gaa, which is critical for the degradation of lysosomal glycogen (500,000 copies/cell); Arsa, which is important for the desulfation of glycolipids (450,000 copies/cell); and Ctsd, the most abundant protease (400,000 copies/cell; Figure 4 upper panel). The four proteins with lowest abundance (Figure 4 lower panel) comprised the three hydrolases Neu4 (2600 copies/cell), Ctss (1565 copies/cell), and Aoah (895 copies/cell), as well as the oligopeptide transporter Abcb9 (2500 copies/cell).

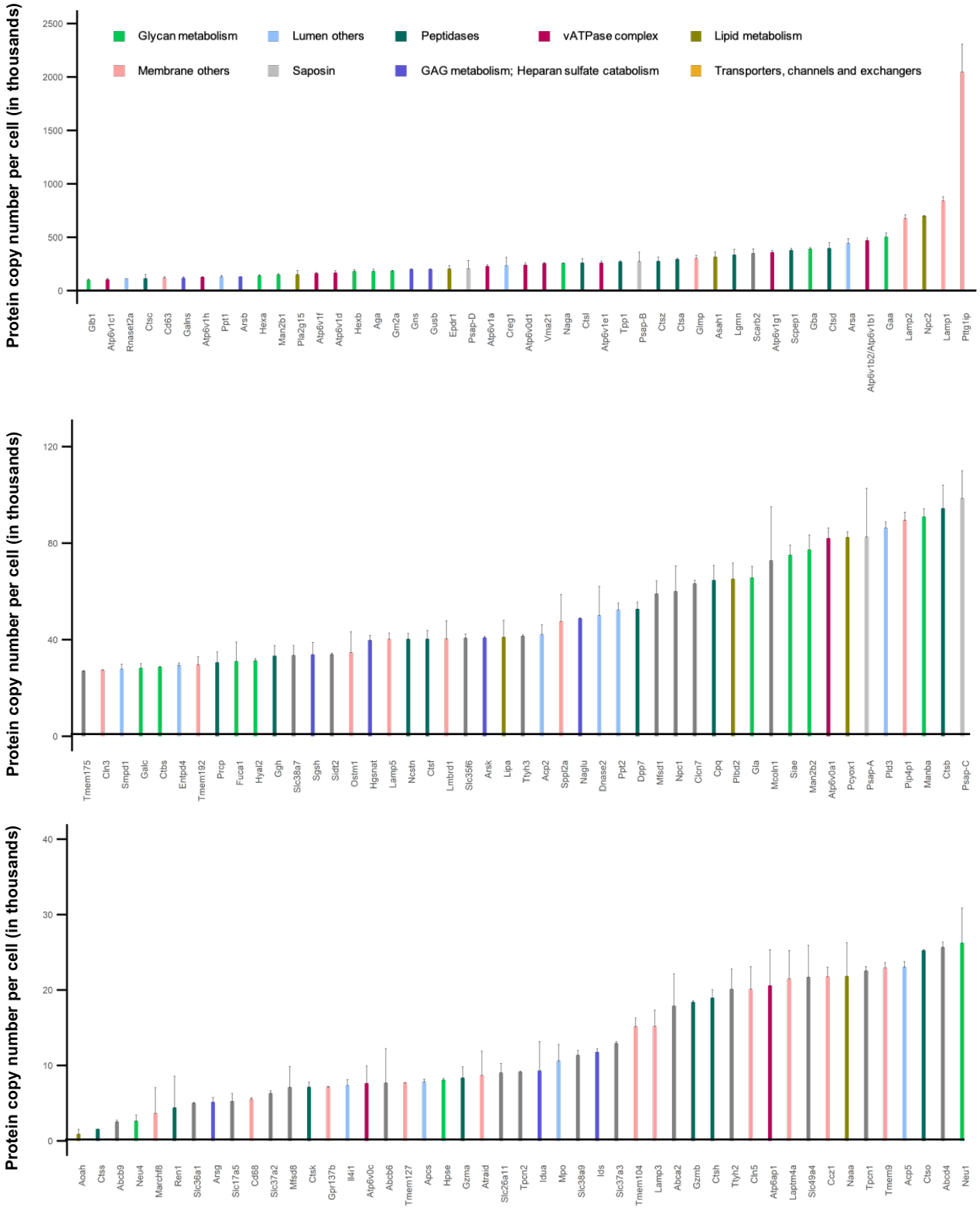


Figure 4: Absolute quantitative analysis of lysosomal proteins in mouse embryonic fibroblasts. Determined protein copy numbers for 141 proteins per cell are shown. Proteins are associated with nine different lysosomal classes /categories. The full range of lysosomal protein copy numbers is divided into three panels: the upper panel displays high abundant proteins, the middle panel shows medium-abundance proteins, and the lower panel highlights low abundant proteins. Median values across biological replicates (n=3) are reported; error bars indicate the robust standard deviation (rSD).

Considering the cumulative protein copy numbers of all lysosomal proteins, a similar distribution of summed abundances for luminal (55.4%) and membrane (44.6%) proteins was observed (Figure 5A). The proteins forming the vATPase complex showed the highest median abundance (~30% of lysosomal membrane proteins) and lowest spread, which is not surprising considering its crucial role for lysosomal function and defined stoichiometry [8]. Among the soluble lysosomal proteins, the group of saposins [306], comprising saposin A, B, C, and D, represents the most abundant protein group based on median abundance (~7% of soluble proteins). With respect to lysosomal hydrolases, the four biggest functional groups account for > 70% of lysosomal luminal protein amount, namely 29% glycosidases (23 proteins), 28% proteases (22 proteins), 10% sulfatases (8 proteins), and 6% lipases (5 proteins). The spread of their absolute protein abundance core distribution, represented by the interquartile range (25th to 75th percentile) for glycosidases, sulfatases, and lipases behaved very similar, while peptidases markedly exceeded the other groups (discrepancy of > 250-fold between highest/lowest abundant members, Figure 5B). For the latter, the sub-group of cathepsins (11 proteins, Figure 5C) was further investigated, which presented with a distribution across the complete dynamic range of peptidase expression levels ranging from 400,000 copies/cell of Ctsd to 1500 copies/cell of Ctss. For most protein classes associated with the lysosomal membrane, a similar spread was observed. However, marked variations in their median expression were noted, with lowest values for transporters, channels, and exchangers (12% of cumulative lysosomal membrane proteins with a class-specific dynamic range of ~140-fold). Within the distinct sub-groups of this class, however, significantly lower dynamic ranges were observed compared to hydrolases, such as 8-fold for ATP-Binding Cassette (ABC) transporters (4 members, Figure 5D), and 10-fold for solute carrier (SLC) transporters (9 members, Figure 5E), respectively.

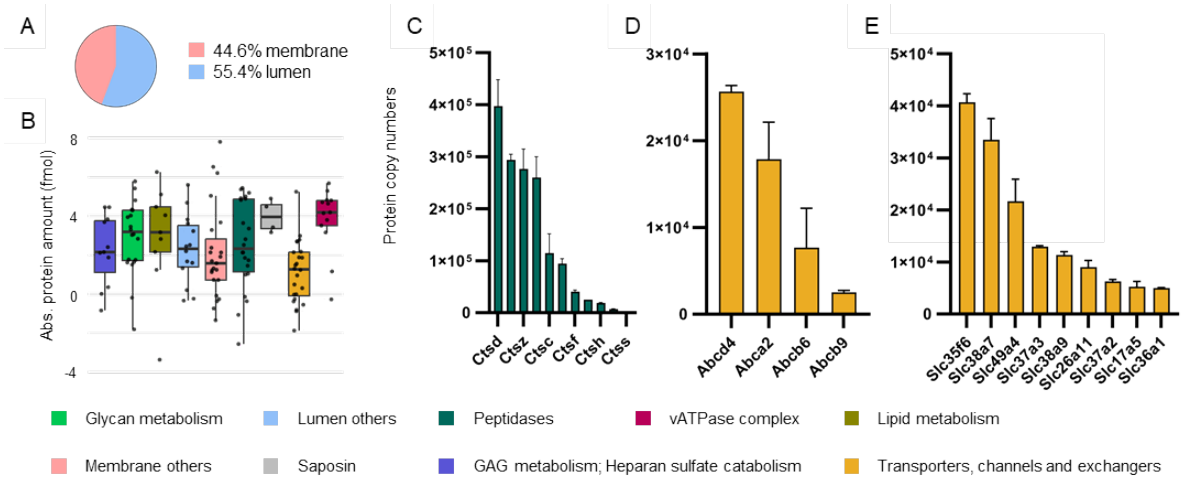


Figure 5: Absolute quantitative analysis of lysosomal proteins in mouse embryonic fibroblasts. A) Distribution of cumulative protein copy numbers. B) Distribution of log2-transformed, absolute protein levels associated with different lysosomal classes/categories. Boxes display the interquartile range (IQR), with each box indicating the median protein expression level for the classes/categories. Whiskers extend to the smallest and largest values within 1.5 times the IQR from the lower and upper quartiles. C, D and E) Copy numbers of lysosomal protein molecules per one MEF cell. Median values across biological replicates (n=3) are reported; error bars indicate the robust standard deviation (rSD).

5.3.4 Determination of lysosomal protein copy numbers from lysosome-enriched fractions

The data presented so far provides information on copy numbers of lysosomal proteins in WCLs but not specifically in lysosomes. To enable the quantitative determination of copy numbers for lysosomal proteins localized within lysosomes, intact lysosomes were enriched using SPIONs (Figure 6A [117]). Resulting lysosome-enriched fractions (LEFs) were analyzed with the MRM assay in the presence of SILLPS (Table S2). As it is impossible to determine the exact number of lysosomes contained in LEFs, and hence to calculate protein copy numbers per organelle analogous to whole cell data, an alternative strategy based on a combination of fluorescence microscopy and protein copy number normalization per lysosome was devised. First, co-staining analyses for Lamp2 (antibody-based) and hydrolytically active Ctsd (SirLysosome [307]) were performed. The number of hydrolytically active terminal lysosomes per MEF cell was determined to be 482 under the given cell culture conditions, with 95% of Lamp2-positive structures containing active Ctsd (Figure 6B). At the same time, no non-lysosomal Lamp2 signal was observed, placing all molecules at this organelle.

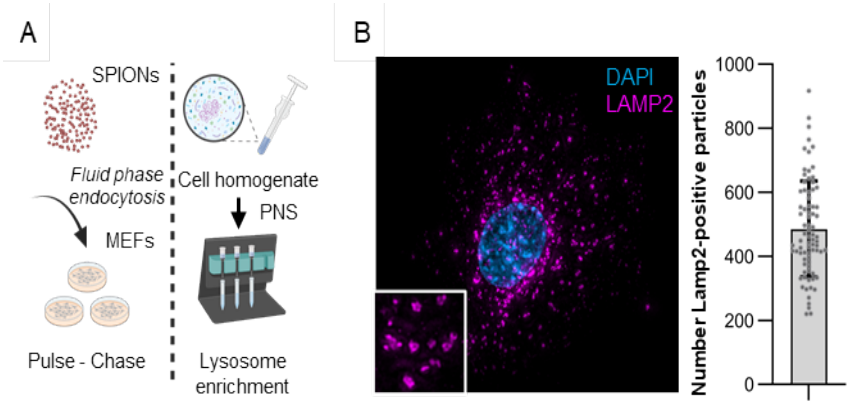


Figure 6: Schematic workflow of intact lysosome isolation and immunofluorescence analysis. A) Lysosomes loaded with paramagnetic particles are retained and isolated from cell homogenate. B) Lysosome-associated membrane protein 2 (Lamp2) immunofluorescence staining in MEF cells (left). Counting of LAMP2-positive markers (n=79, right).

Based on the SILLPS-derived data, ~1400 Lamp2 molecules could be located to each lysosome assuming uniform distribution. This value was used to determine that one LEF from two 10 cm plates of MEFs contains ~7.6x10^7 lysosomes, enabling the calculation of individual protein copy numbers per lysosome both in WCLs and LEFs (Table S2, see Material and methods). For LEFs, a similar dynamic range of individual protein abundance was observed as for the WCL with marked differences for individual proteins (Figure S6). The most abundant proteins in LEFs were saposins B and D (8,400 and 2,500 copies/lysosome, respectively), Lamp2 (1,400 copies/lysosome), Cd63 (1,800 copies/lysosome), and Pttg1ip (1,700 copies/lysosome). This pattern was consistent with the expression of Lamp2 and Pttg1ip observed in WCLs, although there was a marked increase in the relative abundance of the saposins and Cd63 (Figure 7 upper panel). In the case of saposin B and D, this observation was related to the fact that, in LEF samples, one specific peptide from each of these proteins exhibited a disproportionately higher abundance compared to the other peptides of

the same protein. This discrepancy in peptide quantification was however not observed in whole-cell lysates, where the abundances of the two peptides from saposin B and saposin D remained relatively consistent. As a result, the elevated abundance of a single peptide led to a higher overall detected protein quantity of saposin B and saposin D in the LEF sample. Further investigation is needed to determine whether such a discrepancy in peptide quantification, as seen for example with the saposins, arises from factors such as differential peptide stability and protease susceptibility, or from other biases inherent to MS-based sample preparation and quantification. Interestingly, also in these analyses the putative lysosomal protein Pttg1ip was among the five most abundant proteins. While its function remains unclear, its localization within the lysosome appears highly likely and confirms previous findings using orthogonal methods [305].

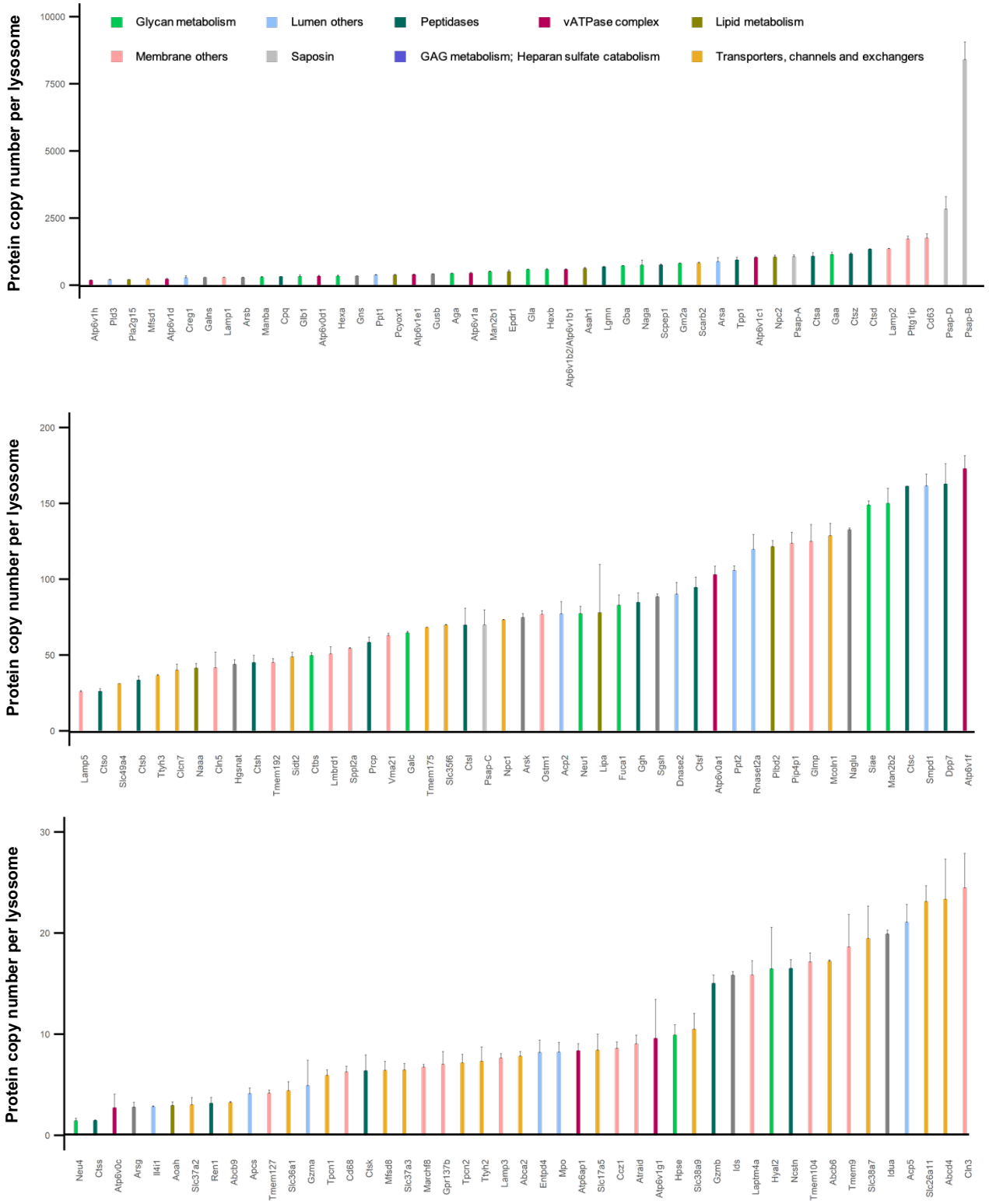


Figure 7: Absolute quantitative analysis of lysosomal proteins in lysosomes-enriched fractions. Determined protein copy numbers for 141 proteins per lysosome are shown. Proteins are associated with nine different lysosomal classes /categories. The full range of lysosomal protein copy numbers is divided into three panels: the upper panel displays the high abundant proteins, the middle panel shows medium-abundance proteins, and the lower panel highlights low abundant proteins. Median values across biological replicates (n=3) are reported; error bars indicate the robust standard deviation (rSD).

5.3.5 Protein distribution analysis identifies differential localization for lysosomal proteins

It is well known that lysosomal proteins are not exclusively lysosomal but are also present in other cellular compartments. For example, while certain luminal lysosomal proteins are almost solely lysosomal, some membrane proteins, such as SLC transporters clearly show also non-lysosomal localization [308]. The comparison of lysosome-normalized protein copy numbers from acquired LEF and WCL data allowed the spatial analysis of cellular distributions for the individual proteins covered by the assay, distinguishing between lysosomal and non-lysosomal localization. As the protein copy numbers of LEF samples, with its solely lysosomal protein content, represents only a fraction of the total lysosomal protein content (comprising protein copies from both, lysosomal and non-lysosomal localization represented by the WCL sample), the difference between both can be attributed to protein copies localizing in non-lysosomal compartments (see Material and methods).

Both datasets were correlated, and the absolute as well as relative proportional distribution for individual proteins was calculated (Table S2). Based on the total number of protein molecules per lysosome, 85% of lysosomal protein molecules were found at the lysosome, whereas the remaining 15% were present in non-lysosomal compartments (Figure 8A). Considering individual proteins, 79 of them showed a primarily lysosomal localization, while for the remaining 62 proteins more than one-third were present at another subcellular location (Figure S7A). The analysis was further extended addressing lysosomal protein classes including luminal hydrolytic, luminal non-hydrolytic and membrane proteins. This revealed that luminal proteins are predominantly located within the lysosome, and clearly higher fractions of non-lysosomal localization for proteins which are located at the lysosomal membrane (Figure 8B).

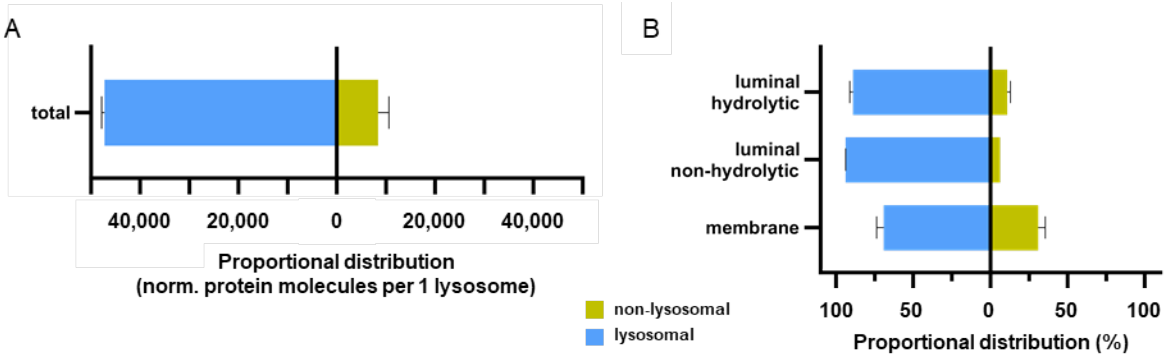


Figure 8: Lysosomal protein distribution analysis. Shown is the absolute and relative proportional distribution of lysosomal protein molecules, normalized to one lysosome, located in/at the lysosome (left) and outside the lysosome (non-lysosomal, right). The sum of individual protein molecules (A) and the sum of individual protein molecules per category was used (B).

However, the further distinction between luminal hydrolytic and luminal non-hydrolytic proteins also showed a difference in their distribution (Figure 8B). Surprisingly, a higher proportion of luminal hydrolytic proteins, compared to luminal non-hydrolytic proteins, was found in non-lysosomal

compartments. In general, these findings indicate that the majority of lysosomal luminal proteins are directly transported to lysosomes, presenting their final destination. On the other hand, membrane proteins may temporarily reside to some extent in the ER and/or Golgi, or permanently localize to entirely different subcellular compartments other than the lysosome.

Broadly, lysosomal membrane proteins can be classified as transporters, channels and exchangers and those with no role in the exchange of metabolites and ions. Of the first class, five proteins showed exclusive lysosomal localization (Figure S7B), from which only one protein each belonged to the two biggest groups, the ABC and SLC transporters (Figure 9A and B). These are the sodium-independent sulfate anion transporter Slc26a11 and the porphyrin transporter Abcb6, which have been shown in other cell lines to also localize to other subcellular compartments but exclusively localize to the lysosome in this MEF data. This could be related to the unique properties of MEF lysosomes, as it was previously demonstrated that their composition differs from that of other commonly used cell lines [56]. With respect to SLC transporters, interestingly, the proteins with the smallest percentage of lysosomal localization (~25%) were the two members of the SLC37 sugar-phosphate exchangers (Slc37a2, Slc37a3, Figure 9B), a SLC protein family initially considered to be ER-resident [309, 310]. Those SLC transporters involved in amino acid transport (Slc38a9 [311], Slc36a1 [312], and Slc38a7 [313]) which have been described as integral components of the lysosomal amino acid-sensing machinery and play a critical role in amino acid sensing by mTORC1 [311, 312], showed a higher degree of lysosomal localization. However, it is notable that more than 50% of each of these transporters was present in non-lysosomal regions of the cell. The lysosomal chloride-proton exchanger Clcn7 and its beta subunit Ostm1 present with unique properties in this context. It was shown previously, that Clcn7 is present in an ER-resident pool and is only transported to lysosomes upon assembly with Ostm1 [314]. In accordance with these findings, 100% of Ostm1 was detected to be localized at lysosomes, while only 31% of Clcn7 was present at the organelle (Figure 9C). A similar picture presented for Mfsd1 and its accessory subunit Glmp, which were recently demonstrated to act as general dipeptide uniporter in lysosomes [295, 315]. While all copies of Mfsd1 localized to lysosomes, only ~70% of Glmp were present there (Figure 9D), potentially reflecting a similar relationship. As structurally relevant components of the lysosomal membrane, the cellular distribution of Lamp1, Lamp2, Lamp3, and Lamp5 was analyzed (Figure 9E). Notably, Lamp2 exhibited full lysosomal localization, whereas only 64% of Lamp1, 27% of Lamp3, and 41% of Lamp5 were localized to the lysosome. This is in line with previous reports placing only a sub-fraction of Lamp1 at degradative lysosomes [52, 316]. Likewise, the vATPase complex presented with a quite heterogenous pattern for the presence of individual subunits with, surprisingly, the highest and lowest rate for two members of the V1 subcomplex (Figure 9F). As the complex exists at lysosomes in a defined stoichiometry, it is possible that such subunits with a high percentage of lysosomal localization present the rate-limiting members for complex formation, while such with a high non-lysosomal fraction are either present in excess or fulfil additional functions in the cell.

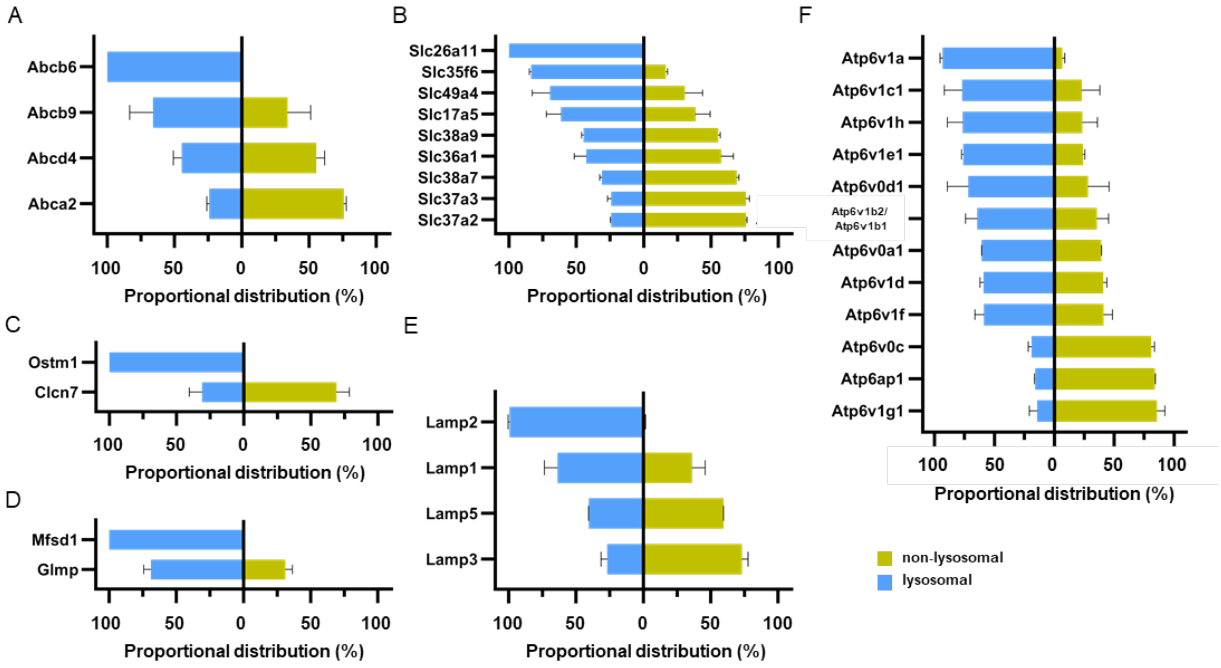


Figure 9: Distribution analysis of lysosomal membrane proteins. Shown is the relative proportional distribution of lysosomal protein molecules, normalized to one lysosome, located in/at the lysosome (left) and outside the lysosome (non-lysosomal, right). Proportional distributions are shown for A) ATP binding cassette transporters (ABCs), B) solute carrier proteins (SLCs), C) osteopetrosis-associated transmembrane protein 1 (Ostm 1) and H(+)/Cl(-) exchange transporter 7 (Clcn 7), D) lysosomal dipeptide transporter (Mfsd 1) and glycosylated lysosomal membrane protein (Glmp), E) lysosome-associated membrane proteins (Lamps) and F) members of the vATPase complex. Abbrev.: ABC: ATP binding cassette transporters, SLC: solute carrier transporter.

The large majority of lysosomal luminal proteins are hydrolases. Their confinement to this subcellular compartment is an essential mechanism to restrict their activity to a secure environment, as mis-localization would result in self-digestion of the cell [317]. Therefore, also the higher degree of lysosomal localization observed for this class of proteins, in comparison to those located at the membrane (Figure 8B), along with the fact that the majority of fully lysosome-localized proteins are found in the lumen, is expected (Figure S7A). Interestingly, with respect to individual classes of hydrolases, such related to glycan metabolism showed the highest degree of lysosomal localization, followed by peptidases, proteins facilitating the degradation of glycosaminoglycans/heparan sulfate and those involved in lipid degradation (Figure S7C, D, E and F).

Peptidases present a distinct group in this context, as all of them catalyze the cleavage of catalytic bonds in a rather unspecific way [318], and hence present with a certain redundancy [319] while other lysosomal hydrolases engage in highly specific degradative reactions. Four of the five highest abundant peptidases (Ctsd, Lgmn, Ctsa, and Ctsz, Figure 4 upper panel) showed exclusive lysosomal localization, while the fifth protein of this group, Scpep1, was located with ~80% at the organelle (Figure S7D). This leads to the majority of peptide hydrolytic activity being concentrated in the lysosome, with more than 60% of total peptidase copy numbers present in this compartment.

The opposite behavior was observed for Ctsk, Ctss, and Ctsb, with only 43% of Ctsk, 24% of Ctss and 16% of Ctsb localized to the lysosome (Figure 10A). Notably, especially Ctsb was shown in numerous studies to exert non-lysosomal functions with involvement in tissue invasion, apoptosis [320], or inflammation [321], which is well in line with a substantial fraction of this protein localized in other subcellular compartments.

Two other hydrolases, detected to be mainly of non-lysosomal localization, are Arsg (17% lysosomal, Figure 10B) and Hyal2 (19% lysosomal, Figure 10C), both involved in catabolism of the large glycosaminoglycans heparan sulfate and hyaluronan, respectively. Arsg is responsible for the degradation of O-sulfated N-sulfoglucosamine residues in heparan sulfate and was demonstrated to localize to pre-lysosomal compartments associating closely with organelle membranes, likely those of the ER [322]. Contradicting reports exist regarding the proteolytic processing of Arsg, with both unprocessed and processed protein forms being associated with hydrolytic activity [322, 323]. For Hyal2, on the other hand, there is an ongoing debate regarding its localization [324-327], which the data clearly identifies as mainly non-lysosomal in MEFs. The fact, that both of these hydrolases, which are involved in the degradation of constituents of the ECM, rank among the least lysosomal-localized proteins, and that they may not require proteolytic activation presenting with extra-lysosomal activity [322, 323], could indicate a function in another subcellular compartment, possible in the maturation of molecules before their transport to the extracellular space.

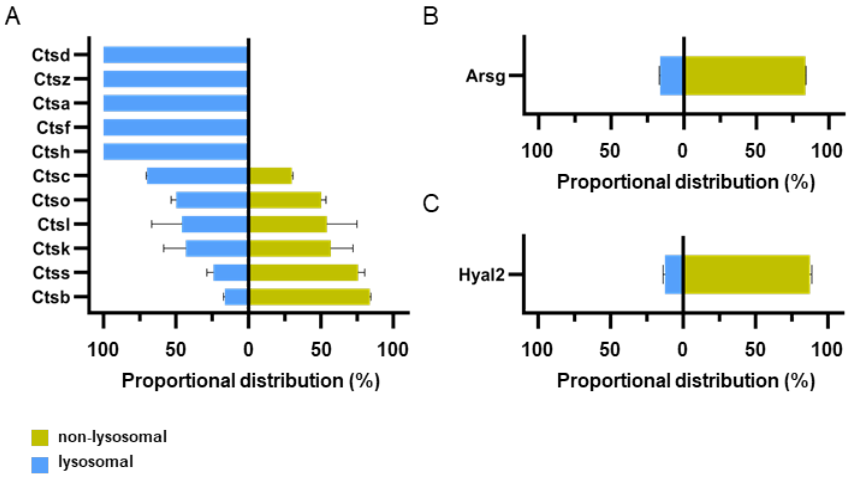


Figure 10: Distribution analysis of lysosomal luminal proteins. Shown is the relative proportional distribution of lysosomal protein molecules, normalized to one lysosome, located in/at the lysosome (left) and outside the lysosome (non-lysosomal, right). Proportional distributions are shown for A) cathepsins, B) arylsulfatase G (Arsg), and C) hyaluronidase 2 (Hyal2).

5.3.6 Cell type-specific lysosomal protein expression patterns show group-specific similarities

As MEF cells, like all other commonly used immortalized cell lines, have been grown in tissue culture for countless passages, it is questionable in how far their proteome is a realistic representation of cells growing in solid tissues. Moreover, it was shown that even within the same cell line grown at different laboratories major proteome differences exist [328]. Primary cells present a more realistic picture and are often the only option to obtain a sufficiently pure population of the cells of interest, as the enrichment of distinct cell types from tissue by FACS or Magnetic-Activated Cell Sorting is often problematic. They are, however, often not available in the quantities needed to enrich lysosomes. In addition, the establishment of lysosome enrichment conditions by Lyso-IP or SPIONs is laborious and requires extensive growth in culture, which may again alter the cellular proteome of the cells [329]. The developed MRM assay presents an ideal solution to investigate such cells. It was therefore applied to primary fibroblasts (as direct comparison to MEFs), as well as three cell types with distinct lysosomal features and most likely differences in lysosomal protein composition. In detail, lung fibroblasts (LFs), osteoblasts (OBs), osteoclasts (OCs), and macrophages (MPs) were investigated, which were either cultured directly from tissue homogenates or generated through differentiation of pluripotent hematopoietic stem cells [330-332].

After determining the lysosomal protein amounts in primary cell types using LC-MRM-MS in combination with SILLPS (Table S3), acquired data were compared to the data generated from MEFs. For MEFs, a markedly lower cumulative amount of lysosomal proteins was observed relative to primary cells, with differences of up to 4-fold. Highest cumulative abundances of lysosomal proteins were found in OBs, OCs, and MPs (Figure 11A). To investigate if differences in abundance are due to a specific portion of the lysosomal proteome, the protein intensity distributions of lysosomal luminal and membrane proteins for each cell line were investigated (Figure 11B), revealing a higher relative contribution of hydrolases in cells with higher lysosomal activity. The observed difference in lysosomal proteome composition was quantitatively determined, revealing a difference in composition from ~60% luminal/~40% membrane proteins in LF and MEF to ~80% luminal/~20% membrane proteins in OB, OC, and MP (Figure S8A). Next, binary comparisons between the different cell types using Pearson correlation were performed to address differences in individual lysosomal proteins. High levels of similarity were observed for the majority of lysosomal proteins between MEFs and LFs, as well as OBs, OCs, and MPs, with pronounced differences evident between the two groups of cells (Figure S8B). Also, unsupervised hierarchical clustering showed similar patterns within these groups and, interestingly, failed to distinguish between MPs and OCs, indicating a high similarity of their lysosomal proteome (Figure 11C). To further investigate the differences between the individual cell types, the summed abundance of individual classes of proteins was compared (Figure 11D, Figure S8C). This revealed that especially the amount of

vATPase complex members, peptidases, and saposins were markedly higher in OBs, OCs, and MPs which is indicative of a higher degradative capacity, and possibly substrate turnover, of these cells. Also, these data showed that the ratio of the vATPase complex to other structural membrane proteins is conserved across primary cell types, implying a similar organization of the lysosomal membrane in these cells, despite their varying luminal content.

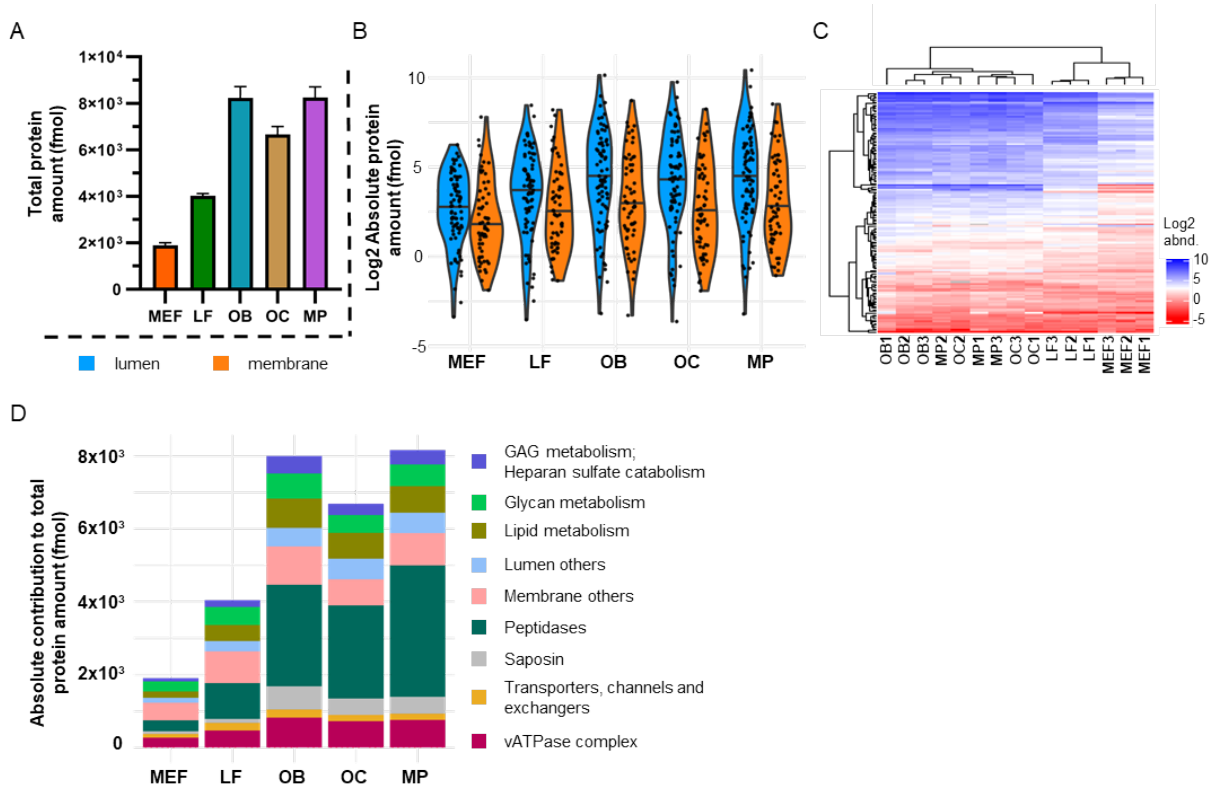


Figure 11: Cell type-specific analysis of lysosomal protein expression. A) Comparison of summed protein abundances across cell types. B) Intensity distribution of luminal and membrane protein abundances. C) Heatmap of hierarchical clustering (complete linkage, Euclidean distance, column- and row-wise clustering) for 141 proteins. Missing values are indicated in grey. D) Absolute contribution of protein classes to the total lysosomal protein abundance, based on summed protein levels per class. For the analysis in Figure 11B and C protein abundance values were log2-transformed. Median expression and robust standard deviation (rSD) are used/indicated in Figure 11A, B and D. Abbrev.: OB: osteoblast, OC: osteoclast, LF: lung fibroblast, MEF: mouse embryonic fibroblast, MP: macrophage.

5.3.7 Cell type-specific dynamics of functionally connected lysosomal protein groups

While the previous analyses enable the identification of protein class-specific differences, they fail to reveal a possible heterogeneity within a respective group of proteins. Therefore, the relative compositions of individual protein groups were further investigated (Figure 12, 13 and 14). With respect to structural membrane proteins, a similar behavior across cell lines for most proteins was observed, with the exception of Lamp1/Lamp2 ratios (Figure 12A). While MEFs expressed both proteins at similar ratios, in primary cells copy numbers of Lamp1 exceeded those of Lamp2 by more than 2-fold (Figure S8D). As the localization of Lamp1 to lysosomes was identified to be well below 100% in MEFs (Figure 9E), and no non-vesicular localization of Lamp1 has been reported while at the same time Lamp1 was further reported to localize to other punctate structures distinct from hydrolytic Ctsd-positive lysosomes [52], this finding suggests either the possibility of Lamp1-only positive lysosomes or Lamp1 localization in an entirely different compartment. For the other major group of membrane proteins, which includes transporters, channels, and exchangers, there were notable discrepancies in individual protein expression between MEFs/LFs and OBs/OCs/MPs (Figure 12B).

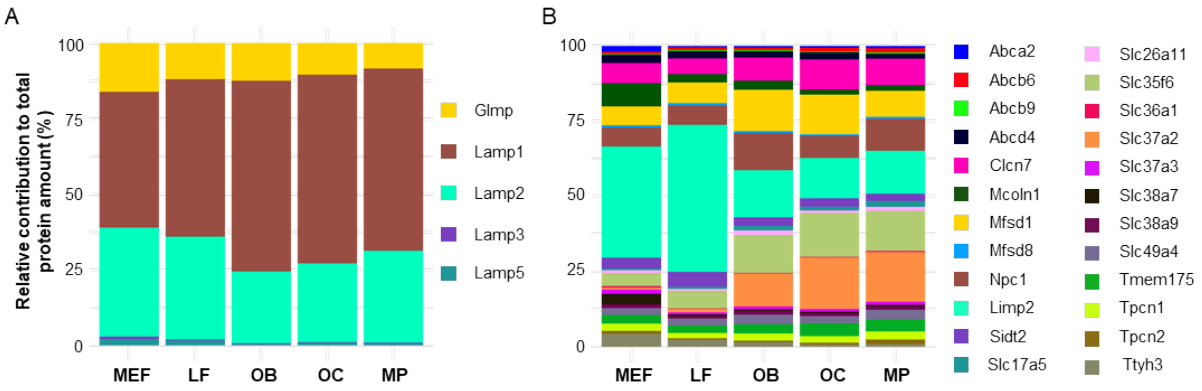


Figure 12: Quantitative composition of classes/categories in individual cell types. Shown is the relative contribution (%) as part of the total protein amount per class/category for A) structural membrane proteins, and B) transporters, channels and exchangers. Abbrev.: OB: osteoblast, OC: osteoclast, LF: lung fibroblast, MEF: mouse embryonic fibroblast, MP: macrophage.

This was especially evident for the cholesterol-trafficking protein Limp2 [333], which constitutes almost 50% of transporters in LFs, but only around 15% in OBs, OCs, and MPs. In contrast, the phosphate-to-glucose 6-phosphate antiporter Slc37a2 exhibited a strong expression in OBs/OCs and MPs (10-20%) compared to MEFs and LFs (<1%). With regard to the total abundance of SLC transporters, Slc37a2, Slc35f6, and Slc38a7 dominated the SLC transporter group in different ratios for the individual cell lines (Figure 13).

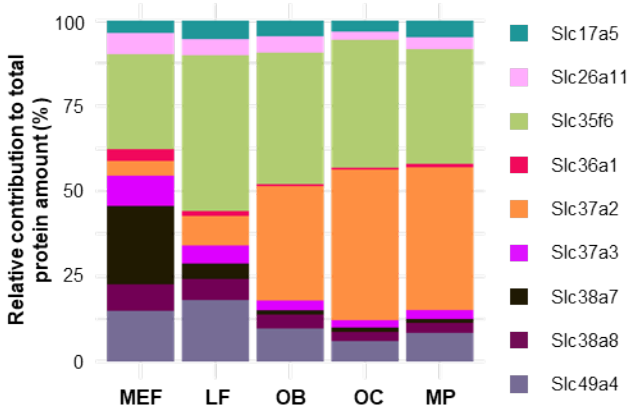


Figure 13: Quantitative composition of classes/categories in individual cell types. Shown is the relative contribution (%) as part of the total protein amount per class/category for solute carrier proteins (SLCs). Abbrev.: OB: osteoblast, OC: osteoclast, LF: lung fibroblast, MEF: mouse embryonic fibroblast, MP: macrophage, SLC: solute carrier protein.

High protein expression levels of the Slc37a2 phosphate/glucose-6-phosphate antiporter [334] were found in OB, OC, and MP, where this transporter constitutes around 35-45% of the total SLC transporter abundance. In contrast, its expression in MEF and LF was considerably lower (4-10%). Expression levels of the sugar transporter Slc35f6 [335], which may be involved in the starvation-induced salvage and cellular reutilization of ribosome-derived nucleosides [336], were comparable across all cell types (30-45%). Functionally relevant is possibly the strong expression of Slc38a7 in MEF cells (25%), as compared to primary cells (≤ 4%). Slc38a7 is a glutamine/asparagine sodium ion symporter that senses lysosomal amino acid content and accordingly impacts the activity of mTORC1 kinase [313, 337]. Given the major role of mTORC1 kinase in regulating cellular metabolism, the metabolic adaptations of immortalized MEF cells - cultivated for extended periods in glutamine-supplemented medium - have likely influenced the expression of Slc38a7. For ABC transporters, a certain heterogeneity in protein expression levels was also observed, although it was not as pronounced as for SLC transporters (Figure S8E). However, it has to be kept in mind that particularly members of the SLC and ABC family are also found in non-lysosomal subcellular locations (Figure S7B). For example, Slc37a2 was the least lysosomal-localized SLC in MEFs (Figure 9B) and was shown previously to be localized at the ER [334]. Since the comparisons of different cell types presented here are based on whole-cell sample observations, differences in protein copy numbers - and consequently, relative expression levels - do not necessarily reflect differences in lysosome-located protein abundance.

Finally, the dynamics of lysosomal hydrolases between the individual cell types were investigated, categorizing them based on the type of substrate group. With respect to the relative summed abundance of individual classes, the MEFs/LFs luminal proteome contained higher levels of glycosidases and sulfatases. OBs, OCs, and MPs contained markedly higher amounts of proteases/peptidases, while lipases were present at similar levels in all cell types (Figure 14A). However, within these individual classes, the relative expression patterns of individual proteins were rather similar, and only individual proteins represented outliers with respect to the class-specific

expression patterns. In the case of glycosidases, for example, Gaa and Gba, which are involved in glycogen degradation [338] and removal of a glucose moiety from the membrane lipid glucocerebroside [339], respectively, these two proteins were markedly more abundant in MEFs and LFs (Figure 14B). In OBs, OCs, and MPs, on the other hand, Gusb, which catalyzes the removal of terminal glucuronic acid from glycosaminoglycans, was very prominent. Among the peptidases, Ctsl was almost exclusively expressed in MEFs, while Ctss was virtually absent in these cells. In contrast, Ctss was expressed in all primary cells, with especially high amounts in OBs, OCs, and MPs (Figure 14C). Ctss has been associated mainly with ECM remodeling and bone/cartilage-related processes. Its deficiency strongly affects tissues, which aligns with the increased expression observed in bone-related cell types in the analysis [340].

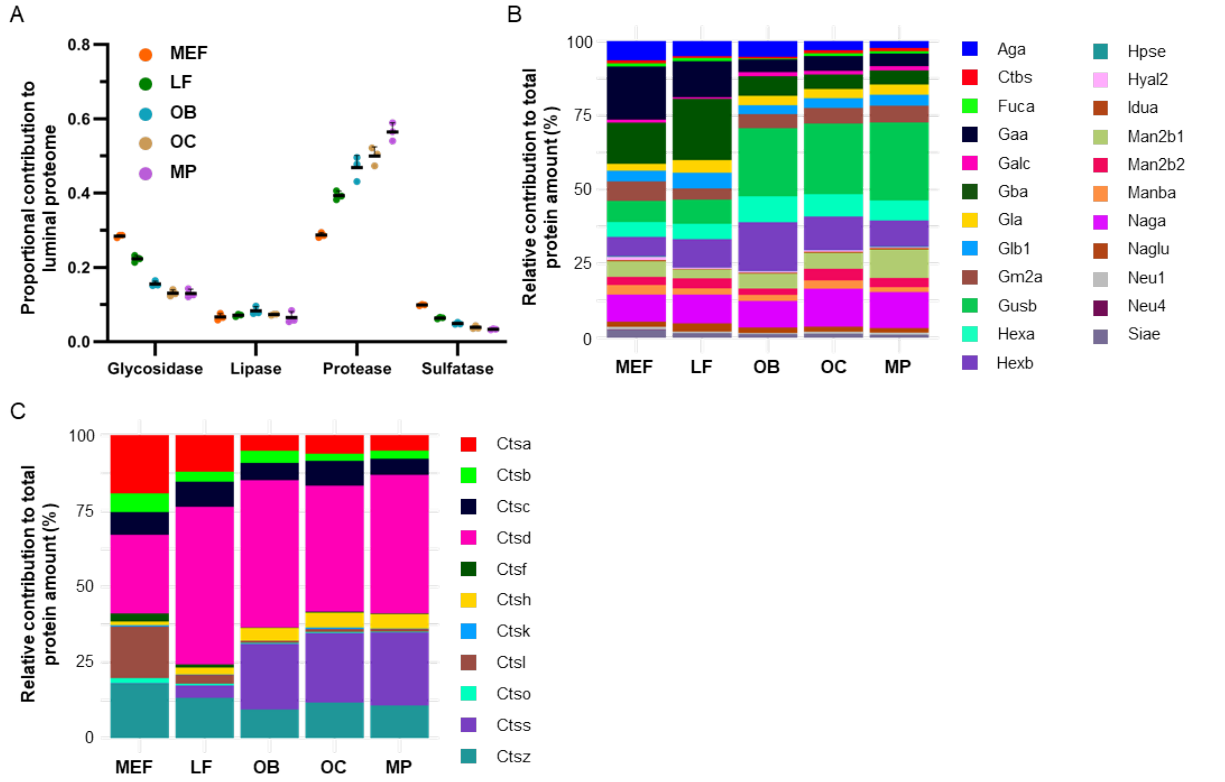


Figure 14: Quantitative composition of classes/categories in individual cell types. Shown is the relative contribution as part of the total protein amount per class/category for A) four hydrolase subtypes as a fraction of the total luminal proteome, B) glycosidases and C) cathepsins. Relative contribution (%) is provided for B) and C). Reported is the average across biological replicates (n=3); error bars indicate the standard deviation (SD). Abbrev.: OB: osteoblast, OC: osteoclast, LF: lung fibroblast, MEF: mouse embryonic fibroblast, MP: macrophage.

5.3.8 Transcriptional and post-transcriptional regulation of lysosomal luminal protein levels

The observed intercellular heterogeneity at the protein level could be due to varying transcription rates in the individual cell types or post-transcriptional processes. It has been shown previously, for example, that Ctsl, Ctsk and Ctsb expression levels are transcriptionally regulated by various transcription factors, e.g. NF-Y, Sp1, Sp3, erythroblast transformation-specific family factors [341, 342] as well as Mitf and TFE3 [343]. To discriminate between these two possibilities, individual qPCR analyses for 51 lysosomal luminal proteins were performed for one fibroblast cell type (LF), one bone derived cell type (OBs) and macrophages (MPs). Datasets were normalized to the respective maximum values within each dataset, and the normalized expression levels were then correlated (Figure 15). Assuming that qPCR efficiencies are comparable for different RNA targets, very similar trends in relative transcript and protein levels were observed among the three cell types, indicating that the observed differences in protein expression are predominantly regulated on a transcriptional level. In the case of Ctsd, for example, high transcript levels resulted in high protein levels. However, there are a few notable exceptions where relative protein levels did not correspond to transcript levels or at least not across all cell types, as seen in the case of Ctsb. Here, Ctsb exhibited high transcript levels in LF and OB but low protein expression. In contrast, MP presented with low Ctsb transcript levels, which resulted in low Ctsb protein expression. Such a finding suggests that Ctsb expression in LF and OB is heavily influenced by post-transcriptional mechanisms. However, other cellular processes, such as pronounced exocytosis of these proteins [344] or a high proteolytic turnover in the lysosome, are also conceivable.

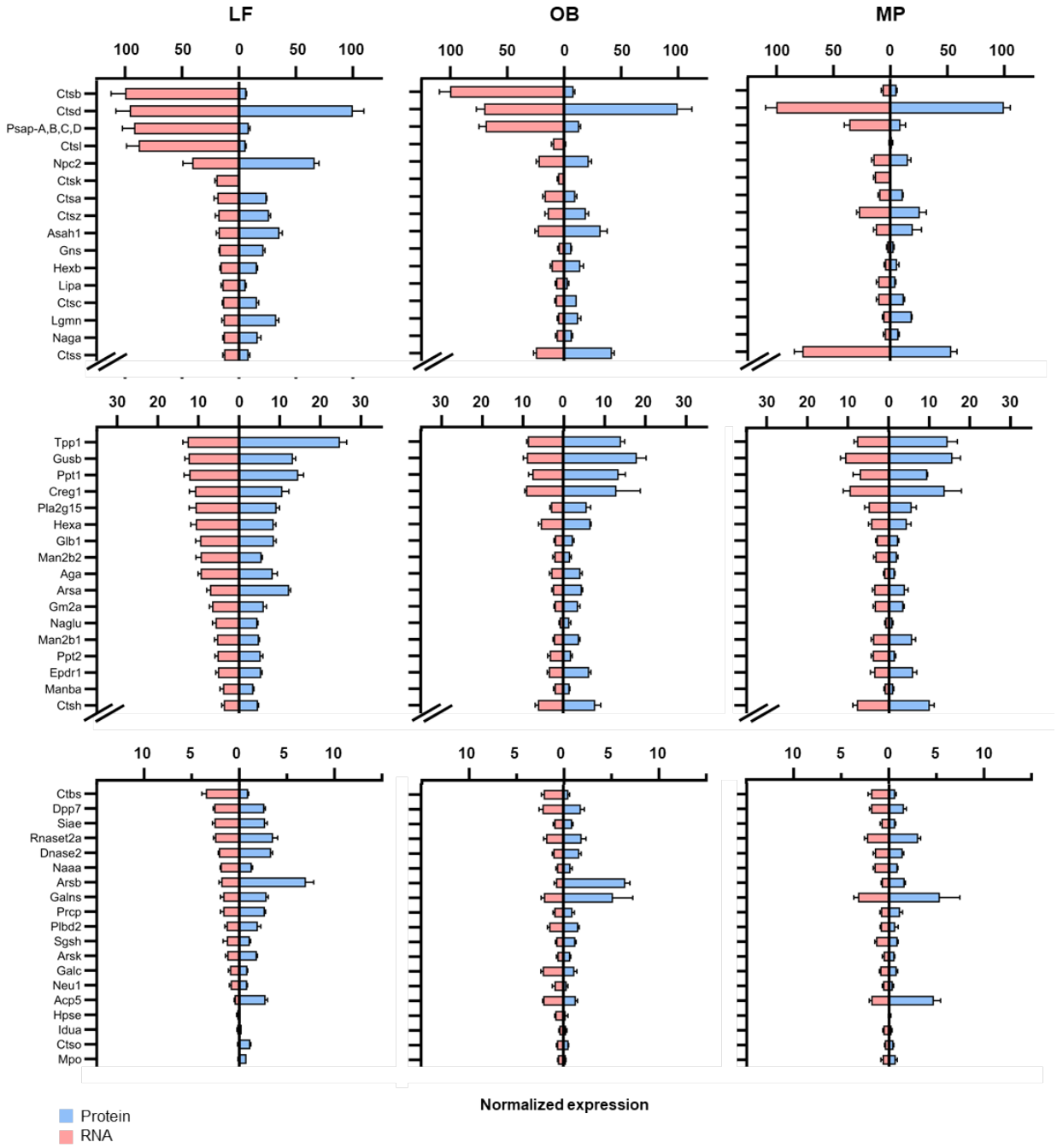


Figure 15: Comparison of RNA and protein expression profiles in primary cells. Mirror plots show normalized RNA and protein expression levels scaled to the dataset's maximum expression. Error bars indicate the scaled standard deviation (SD). Abbrev.: LF: lung fibroblast, OB: osteoblast, MP: macrophage.

5.4 Conclusion

For the unbiased and enrichment-free absolute quantification of lysosomal proteins from whole cell and tissue lysates, a targeted MS platform was established, using a combination of MRM-MS and SILLPSs. Therefore, a set of 12 stable isotope-labeled artificial protein standards, applying the QconCAT strategy was designed, expressed and digested. The generated and absolute quantified peptide standard comprised 414 peptides, serving as surrogates in the quantification of 144 lysosomal proteins. Following standard characterization and validation of efficient peptide release, a tailored MRM assay for highly reproducible and sensitive quantification was developed. This robust and universally applicable strategy enabled the absolute quantification of up to 143 lysosomal proteins based on 388 proteotypic surrogate peptides.

Applied to whole cell lysate of MEFs and four primary cell types as well as LEFs, the strategy provided absolute protein copy numbers allowing for the first time the absolute quantitative investigation of the lysosomal proteome composition. The analysis of lysosomal protein copy numbers in MEFs revealed a dynamic range spanning over three orders of magnitude, with Lamp1 and Lamp2 among the most abundant proteins, comprising ~22% of lysosomal membrane protein biomass. Other highly abundant proteins included proteins forming the vATPase complex, which is essential for lysosomal lumen acidification. The class of lysosomal hydrolytic proteins was dominated by the four major functional groups: glycosidases, proteases, sulfatases, and lipases, which account cumulative for > 70% of the luminal protein amount. Among these functional groups, proteases/peptidases exhibited a wide range of protein abundances, with a >250-fold discrepancy between the highest and lowest abundant members of the group. Further, the correlation of protein copy numbers from MEF whole cell and LEFs, combined with fluorescence microscopy and protein copy number normalization, enabled the investigation of spatial protein distribution between the lysosomal and non-lysosomal compartment. While ~85% of the total lysosomal protein molecules were located in or at the organelle, group-specific distribution patterns were identified localizing substantial proportions of membrane-bound proteins in non-lysosomal compartments. Examples of these observations include, among others, members of the SLC transporter family as well as the lysosomal chloride-proton exchanger Clcn7. But also, for luminal hydrolytically active proteins, for which as expected the majority presented with a lysosomal localization, examples of non-lysosomal localization were detected. Examples include the peptidases Ctsk, Ctss and Ctsb, as well as the sulfatase Arsg and the glycosidase Hyal2. In general, such findings imply the presence of additional lysosomal protein pools at subcellular locations other than the lysosome, or at least the retention of lysosomal proteins in non-lysosomal compartments. In case of the luminal, hydrolytically active proteins, the possibility of lysosomal hydrolase function and involvement in cellular processes beyond the lysosome itself should be considered.

Further, the established approach was also applied to the investigation of primary cells. Primary cells reflect the *in vivo* conditions of their tissue of origin more accurately than immortalized cell lines such as MEFs, thereby providing a more realistic system for studying lysosomal function and the proteome. Their limited availability had so far largely excluded primary cells from enrichment-based MS analysis strategies. The enrichment-free, targeted MS platform enabled the analysis of whole cell lysates from LFs, OBs, OCs and MPs. Determined lysosomal protein amounts were compared with the data from the initial MEF experiments. Across the different cell types investigated, primary cells had a markedly increased total lysosomal protein amount compared to MEFs, with up to a 4-fold increase observed in MPs, OBs and OCs. Differences in lysosomal protein amounts were most likely related to increased lysosomal activity, as a high percentage of the overrepresented proteins were hydrolases, i.e. luminal proteins within the lysosomal proteome in these cells. In general, the global proteome profiling showed strong similarities in lysosomal protein expression of MEFs and LFs, as well as among OBs, OCs, and MPs, with distinct differences between the two groups. Comparison of summed protein abundances for individual classes between the five cell types revealed elevated amounts of vATPase complex members, peptidases and saposins for OBs, OCs and MPs compared to MEFs and LFs.

This further suggests a higher catabolic capacity in these primary cell types and underscores their functional specialization, which is accompanied by distinct adaptations in lysosomal protein class expression. At the same time, the overall relative composition of individual protein groups and classes, as seen for example for members of the vATPase complex as well as structural membrane proteins, remained quite stable across the cell types investigated. Here, only variations in the expression of selected proteins were discovered in multiple cases. Interestingly, individual cell types showed unique expression dynamics for lysosomal hydrolases when categorized by their substrate groups. For instance, the proportion of glycosidases and sulfatases as part of the luminal proteome was higher in MEFs and LFs, while OBs, OCs and MPs presented with higher amounts of peptidases. The correlation of protein and transcript levels for 51 luminal proteins across three primary cell types further suggested that the observed differences in cell type-specific protein expression are primarily governed by direct transcriptional regulation.

6. Outlook and Future Directions

In the presented work the absolute quantification of lysosomal proteins in a comprehensive manner using targeted MS was established. The combination of QconCAT-derived peptide standards and MRM-MS enabled the direct absolute quantification of low-abundant lysosomal proteins from mouse cell lines and, in principle, of mouse samples irrespective of their origin. This eliminated the need for lysosome enrichment or isolation, allowing for an unbiased, quantitative study of lysosomal proteins across cell types, even with limited input. As an initial study, over 100 lysosomal proteins were systematically mapped across different cell types, providing protein copy number data.

The data and findings presented in this study showcase the establishment of a technological platform comprising MS-data acquisition as well as data analysis and visualization. This can serve as a starting point for a deeper understanding of lysosomal molecular biology, and lays the foundation for future investigations. The next research phase should encompass the direct analysis of tissue/organ samples, as these represent the more important biological matrix in future studies and animal-based disease models are essential for the investigation of lysosomal storage disorders (LSDs). As it is highly plausible that different tissue types express lysosomal proteomes adapted to their specific functional roles, it would be intriguing to quantitatively map these tissue-specific lysosomal proteomes and uncover their unique features. Such knowledge could also shed light on the observed characteristic organ-specific phenotypes of LSDs. Following this rather broad exploration, the investigation of animal-based disease models for lysosomal storage diseases should be considered. The study of animal models will provide critical insights into the molecular mechanisms of the diseases, their broader impact on affected lysosomal pathways and/or the lysosome, and should help to identify potential therapeutic targets and strategies. Additionally, for future research, the broader implications of lysosomal changes on overall cell physiology and, consequently, the cell’s proteome should be investigated.

7. References

1. Fujiwara, Y., K. Wada, and T. Kabuta, *Lysosomal degradation of intracellular nucleic acids-multiple autophagic pathways.* J Biochem, 2017. **161**(2): p. 145-154.
2. Settembre, C., et al., *Signals from the lysosome: a control centre for cellular clearance and energy metabolism.* Nat Rev Mol Cell Biol, 2013. **14**(5): p. 283-96.
3. Goebel, T., et al., *Proteaphagy in Mammalian Cells Can Function Independent of ATG5/ATG7.* Mol Cell Proteomics, 2020. **19**(7): p. 1120-1131.
4. Uoselis, L., T.N. Nguyen, and M. Lazarou, *Mitochondrial degradation: Mitophagy and beyond.* Mol Cell, 2023. **83**(19): p. 3404-3420.
5. Rosales, C. and E. Uribe-Querol, *Phagocytosis: A Fundamental Process in Immunity.* Biomed Res Int, 2017. **2017**: p. 9042851.
6. Luzio, J.P., P.R. Pryor, and N.A. Bright, *Lysosomes: fusion and function.* Nat Rev Mol Cell Biol, 2007. **8**(8): p. 622-32.
7. Billman, G.E., *Homeostasis: The Underappreciated and Far Too Often Ignored Central Organizing Principle of Physiology.* Front Physiol, 2020. **11**: p. 200.
8. Abbas, Y.M., et al., *Structure of V-ATPase from the mammalian brain.* Science, 2020. **367**(6483): p. 1240-1246.
9. Muthukottiappan, P. and D. Winter, *A proteomic view on lysosomes.* Mol Omics, 2021. **17**(6): p. 842-859.
10. Tancini, B., et al., *Lysosomal Exocytosis: The Extracellular Role of an Intracellular Organelle.* Membranes, 2020. **10**(12): p. 406.
11. Lim, C.Y. and R. Zoncu, *The lysosome as a command-and-control center for cellular metabolism.* J Cell Biol, 2016. **214**(6): p. 653-64.
12. De Duve, C., et al., *Intracellular distribution patterns of enzymes in rat-liver tissue.* Biochem J, 1955. **60**(4): p. 604-17.
13. Sancak, Y., et al., *Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids.* Cell, 2010. **141**(2): p. 290-303.
14. Sardiello, M., et al., *A gene network regulating lysosomal biogenesis and function.* Science, 2009. **325**(5939): p. 473-7.
15. Roczniak-Ferguson, A., et al., *The transcription factor TFEB links mTORC1 signaling to transcriptional control of lysosome homeostasis.* Sci Signal, 2012. **5**(228): p. ra42.
16. Garg, S., et al., *Lysosomal trafficking, antigen presentation, and microbial killing are controlled by the Arf-like GTPase Arl8b.* Immunity, 2011. **35**(2): p. 182-93.
17. Reddy, A., E.V. Caler, and N.W. Andrews, *Plasma membrane repair is mediated by Ca(2+)-regulated exocytosis of lysosomes.* Cell, 2001. **106**(2): p. 157-69.
18. Buratta, S., et al., *Lysosomal Exocytosis, Exosome Release and Secretory Autophagy: The Autophagic- and Endo-Lysosomal Systems Go Extracellular.* Int J Mol Sci, 2020. **21**(7): p. 2576.
19. Pu, J., et al., *Mechanisms and functions of lysosome positioning.* J Cell Sci, 2016. **129**(23): p. 4329-4339.
20. Meng, Y., et al., *Cholesterol Handling in Lysosomes and Beyond.* Trends Cell Biol, 2020. **30**(6): p. 452-466.
21. Hao, F., et al., *Rheb localized on the Golgi membrane activates lysosome-localized mTORC1 at the Golgi-lysosome contact site.* J Cell Sci, 2018. **131**(3): p. jcs208017.
22. Zhao, Y.G. and H. Zhang, *Autophagosome maturation: An epic journey from the ER to lysosomes.* J Cell Biol, 2019. **218**(3): p. 757-770.
23. Wong, Y.C., D. Ysselstein, and D. Krainc, *Mitochondria-lysosome contacts regulate mitochondrial fission via RAB7 GTP hydrolysis.* Nature, 2018. **554**(7692): p. 382-386.
24. Yagi, M., et al., *Mitochondrial translation deficiency impairs NAD(+) -mediated lysosomal acidification.* Embo j, 2021. **40**(8): p. e105268.
25. Zoncu, R., et al., *mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H(+)-ATPase.* Science, 2011. **334**(6056): p. 678-83.
26. Bautista, S.J., et al., *mTOR complex 1 controls the nuclear localization and function of glycogen synthase kinase 3β.* J Biol Chem, 2018. **293**(38): p. 14723-14739.
27. Zhang, C.S., et al., *The lysosomal v-ATPase-Ragulator complex is a common activator for AMPK and mTORC1, acting as a switch between catabolism and anabolism.* Cell Metab, 2014. **20**(3): p. 526-40.

28. Takahara, T., et al., *Amino acid-dependent control of mTORC1 signaling: a variety of regulatory modes.* J Biomed Sci, 2020. **27**(1): p. 87.
29. Goul, C., R. Peruzzo, and R. Zoncu, *The molecular basis of nutrient sensing and signalling by mTORC1 in metabolism regulation and disease.* Nat Rev Mol Cell Biol, 2023. **24**(12): p. 857-875.
30. Perera, R.M., C. Di Malta, and A. Ballabio, *MiT/TFE Family of Transcription Factors, Lysosomes, and Cancer.* Annu Rev Cancer Biol, 2019. **3**: p. 203-222.
31. Napolitano, G. and A. Ballabio, *TFEB at a glance.* J Cell Sci, 2016. **129**(13): p. 2475-81.
32. Palmieri, M., et al., *Characterization of the CLEAR network reveals an integrated control of cellular clearance pathways.* Hum Mol Genet, 2011. **20**(19): p. 3852-66.
33. Takla, M., S. Keshri, and D.C. Rubinsztein, *The post-translational regulation of transcription factor EB (TFEB) in health and disease.* EMBO Rep, 2023. **24**(11): p. e57574.
34. Eisenstein, M., *Myriad maladies.* Nature, 2016. **537**(7621): p. S146-7.
35. Platt, F.M., et al., *Lysosomal storage diseases.* Nat Rev Dis Primers, 2018. **4**(1): p. 27.
36. Harris, H. and D.C. Rubinsztein, *Control of autophagy as a therapy for neurodegenerative disease.* Nat Rev Neurol, 2011. **8**(2): p. 108-17.
37. Lee, J.H., et al., *Lysosomal proteolysis and autophagy require presenilin 1 and are disrupted by Alzheimer-related PS1 mutations.* Cell, 2010. **141**(7): p. 1146-58.
38. Robak, L.A., et al., *Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease.* Brain, 2017. **140**(12): p. 3191-3203.
39. Davidson, S.M. and M.G. Vander Heiden, *Critical Functions of the Lysosome in Cancer Biology.* Annu Rev Pharmacol Toxicol, 2017. **57**: p. 481-507.
40. Fernández-Pereira, C., et al., *Therapeutic Approaches in Lysosomal Storage Diseases.* Biomolecules, 2021. **11**(12): p. 1775.
41. van der Ploeg, A.T. and A.J. Reuser, *Pompe's disease.* Lancet, 2008. **372**(9646): p. 1342-53.
42. Roach, P.J., et al., *Glycogen and its metabolism: some new developments and old themes.* Biochem J, 2012. **441**(3): p. 763-87.
43. Roig-Zamboni, V., et al., *Structure of human lysosomal acid α-glucosidase-a guide for the treatment of Pompe disease.* Nat Commun, 2017. **8**(1): p. 1111.
44. Fuller, D.D., et al., *The respiratory neuromuscular system in Pompe disease.* Respir Physiol Neurobiol, 2013. **189**(2): p. 241-9.
45. Falk, D.J., et al., *Peripheral nerve and neuromuscular junction pathology in Pompe disease.* Hum Mol Genet, 2015. **24**(3): p. 625-36.
46. Lim, J.A., et al., *Pompe disease: Shared and unshared features of lysosomal storage disorders.* Rare Dis, 2015. **3**(1): p. e1068978.
47. Byers, S., et al., *Glycosaminoglycan accumulation and excretion in the mucopolysaccharidoses: characterization and basis of a diagnostic test for MPS.* Mol Genet Metab, 1998. **65**(4): p. 282-90.
48. Benetó, N., et al., *Sanfilippo Syndrome: Molecular Basis, Disease Models and Therapeutic Approaches.* Int J Mol Sci, 2020. **21**(21): p. 7819.
49. Kowalewski, B., et al., *Arylsulfatase G inactivation causes loss of heparan sulfate 3-O-sulfatase activity and mucopolysaccharidosis in mice.* Proc Natl Acad Sci U S A, 2012. **109**(26): p. 10310-5.
50. Seker Yilmaz, B., et al., *Novel therapies for mucopolysaccharidosis type III.* J Inherit Metab Dis, 2021. **44**(1): p. 129-147.
51. Lavery, C., C.J. Hendriksz, and S.A. Jones, *Mortality in patients with Sanfilippo syndrome.* Orphanet J Rare Dis, 2017. **12**(1): p. 168.
52. Cheng, X.T., et al., *Characterization of LAMP1-labeled nondegradative lysosomal and endocytic compartments in neurons.* J Cell Biol, 2018. **217**(9): p. 3127-3139.
53. Smith, L.M., et al., *The Human Proteoform Project: Defining the human proteome.* Sci Adv, 2021. **7**(46): p. eabk0734.
54. Muntel, J., et al., *Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy.* Mol Omics, 2019. **15**(5): p. 348-360.
55. Messner, C.B., et al., *Ultra-fast proteomics with Scanning SWATH.* Nat Biotechnol, 2021. **39**(7): p. 846-854.
56. Akter, F., et al., *Multi-Cell Line Analysis of Lysosomal Proteomes Reveals Unique Features and Novel Lysosomal Proteins.* Mol Cell Proteomics, 2023. **22**(3): p. 100509.

57. Čaval, T., et al., *Targeted Analysis of Lysosomal Directed Proteins and Their Sites of Mannose-6-phosphate Modification.* Mol Cell Proteomics, 2019. **18**(1): p. 16-27.
58. Thelen, M., et al., *SILAC-Based Comparative Proteomic Analysis of Lysosomes from Mammalian Cells Using LC-MS/MS.* Methods Mol Biol, 2017. **1594**: p. 1-18.
59. Futai, M., et al., *Vacuolar-type ATPase: A proton pump to lysosomal trafficking.* Proc Jpn Acad Ser B Phys Biol Sci, 2019. **95**(6): p. 261-277.
60. Kim, D.H., et al., *mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery.* Cell, 2002. **110**(2): p. 163-75.
61. Laplante, M. and D.M. Sabatini, *mTOR signaling in growth control and disease.* Cell, 2012. **149**(2): p. 274-93.
62. Cabukusta, B. and J. Neefjes, *Mechanisms of lysosomal positioning and movement.* Traffic, 2018. **19**(10): p. 761-769.
63. Balderhaar, H.J. and C. Ungermann, *CORVET and HOPS tethering complexes - coordinators of endosome and lysosome fusion.* J Cell Sci, 2013. **126**(Pt 6): p. 1307-16.
64. Skowyra, M.L., et al., *Triggered recruitment of ESCRT machinery promotes endolysosomal repair.* Science, 2018. **360**(6384): p. aar5078.
65. Eliuk, S. and A. Makarov, *Evolution of Orbitrap Mass Spectrometry Instrumentation.* Annu Rev Anal Chem, 2015. **8**: p. 61-80.
66. Meier, F., et al., *Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer.* Mol Cell Proteomics, 2018. **17**(12): p. 2534-2545.
67. Tsou, C.C., et al., *DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics.* Nat Methods, 2015. **12**(3): p. 258-64.
68. Gessulat, S., et al., *Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning.* Nat Methods, 2019. **16**(6): p. 509-518.
69. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.
70. Sinha, A. and M. Mann, *A beginner's guide to mass spectrometry–based proteomics*, in *The Biochemist*. 2020: London. p. 64–69.
71. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.
72. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.* Anal Chem, 1988. **60**(20): p. 2299-301.
73. Domon, B. and R. Aebersold, *Mass spectrometry and protein analysis.* Science, 2006. **312**(5771): p. 212-7.
74. Steen, H. and M. Mann, *The ABC's (and XYZ's) of peptide sequencing.* Nat Rev Mol Cell Biol, 2004. **5**(9): p. 699-711.
75. Harper, J.W. and E.J. Bennett, *Proteome complexity and the forces that drive proteome imbalance.* Nature, 2016. **537**(7620): p. 328-38.
76. Smith, L.M. and N.L. Kelleher, *Proteoform: a single term describing protein complexity.* Nat Methods, 2013. **10**(3): p. 186-7.
77. Toby, T.K., L. Fornelli, and N.L. Kelleher, *Progress in Top-Down Proteomics and the Analysis of Proteoforms.* Annu Rev Anal Chem, 2016. **9**(1): p. 499-519.
78. Aebersold, R. and M. Mann, *Mass-spectrometric exploration of proteome structure and function.* Nature, 2016. **537**(7620): p. 347-55.
79. Olsen, J.V., S.E. Ong, and M. Mann, *Trypsin cleaves exclusively C-terminal to arginine and lysine residues.* Mol Cell Proteomics, 2004. **3**(6): p. 608-14.
80. Altelaar, A.F., J. Munoz, and A.J. Heck, *Next-generation proteomics: towards an integrative view of proteome dynamics.* Nat Rev Genet, 2013. **14**(1): p. 35-48.
81. Jafari, M., et al., *Comparison of in-gel protein separation techniques commonly used for fractionation in mass spectrometry-based proteomic profiling.* Electrophoresis, 2012. **33**(16): p. 2516-26.
82. Moreda-Piñeiro, A., N. García-Otero, and P. Bermejo-Barrera, *A review on preparative and semi-preparative offgel electrophoresis for multidimensional protein/peptide assessment.* Anal Chim Acta, 2014. **836**: p. 1-17.
83. Manadas, B., et al., *Peptide fractionation in proteomics approaches.* Expert Rev Proteomics, 2010. **7**(5): p. 655-63.
84. Cao, X., et al., *Evaluation of Spin Columns for Human Plasma Depletion to Facilitate MS-Based Proteomics Analysis of Plasma.* J Proteome Res, 2021. **20**(9): p. 4610-4620.

85. Günther, R., et al., *Depletion of highly abundant proteins from human cerebrospinal fluid: a cautionary note.* Mol Neurodegener, 2015. **10**: p. 53.
86. Wang, Q., et al., *Chemical Depletion of Histidine-Containing Peptides Allows Identification of More Low-Abundance Methylation Sites from Proteome Samples.* J Proteome Res, 2021. **20**(5): p. 2497-2505.
87. Loo, J.A., H.R. Udseth, and R.D. Smith, *Peptide and protein analysis by electrospray ionization-mass spectrometry and capillary electrophoresis-mass spectrometry.* Anal Biochem, 1989. **179**(2): p. 404-12.
88. Koopmans, F., et al., *Comparative Analyses of Data Independent Acquisition Mass Spectrometric Approaches: DIA, WiSIM-DIA, and Untargeted DIA.* Proteomics, 2018. **18**(1): p. 1700304.
89. Tabb, D.L., et al., *Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry.* J Proteome Res, 2010. **9**(2): p. 761-76.
90. Gillet, L.C., et al., *Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis.* Mol Cell Proteomics, 2012. **11**(6): p. O111.016717.
91. Ludwig, C., et al., *Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial.* Mol Syst Biol, 2018. **14**(8): p. e8126.
92. Hu, A., W.S. Noble, and A. Wolf-Yadlin, *Technical advances in proteomics: new developments in data-independent acquisition.* F1000Res, 2016. **5**: p. 419.
93. Barkovits, K., et al., *Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition.* Mol Cell Proteomics, 2020. **19**(1): p. 181-197.
94. Kleinnijenhuis, A.J., et al., *Analysis of histidine phosphorylation using tandem MS and ion-electron reactions.* Anal Chem, 2007. **79**(19): p. 7450-6.
95. Reiding, K.R., et al., *The benefits of hybrid fragmentation methods for glycoproteomics.* TrAC Trends in Analytical Chemistry, 2018. **108**: p. 260-268.
96. Roepstorff, P. and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass spectra of peptides.* Biomed Mass Spectrom, 1984. **11**(11): p. 601.
97. Wells, J.M. and S.A. McLuckey, *Collision-induced dissociation (CID) of peptides and proteins.* Methods Enzymol, 2005. **402**: p. 148-85.
98. Marquioni, V., F.M. Franco Nunes, and M.T. Marques Novo-Mansur, *Protein Identification by Database Searching of Mass Spectrometry Data in the Teaching of Proteomics.* Journal of Chemical Education, 2021. **98**(3): p. 812-823.
99. Standing, K.G., *Peptide and protein de novo sequencing by mass spectrometry.* Curr Opin Struct Biol, 2003. **13**(5): p. 595-601.
100. Nesvizhskii, A.I., *Protein identification by tandem mass spectrometry and sequence database searching.* Methods Mol Biol, 2007. **367**: p. 87-119.
101. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.
102. Deutsch, E.W., *Tandem mass spectrometry spectral libraries and library searching.* Methods Mol Biol, 2011. **696**: p. 225-32.
103. Sinitcyn, P., et al., *MaxDIA enables library-based and library-free data-independent acquisition proteomics.* Nat Biotechnol, 2021. **39**(12): p. 1563-1573.
104. Demichev, V., et al., *DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput.* Nat Methods, 2020. **17**(1): p. 41-44.
105. Valm, A.M., et al., *Applying systems-level spectral imaging and analysis to reveal the organelle interactome.* Nature, 2017. **546**(7656): p. 162-167.
106. Ba, Q., et al., *Whole-Cell Scale Dynamic Organization of Lysosomes Revealed by Spatial Statistical Analysis.* Cell Rep, 2018. **23**(12): p. 3591-3606.
107. Itzhak, D.N., et al., *A Mass Spectrometry-Based Approach for Mapping Protein Subcellular Localization Reveals the Spatial Proteome of Mouse Primary Neurons.* Cell Rep, 2017. **20**(11): p. 2706-2718.
108. Itzhak, D.N., et al., *Global, quantitative and dynamic mapping of protein subcellular localization.* Elife, 2016. **5**: p. e16950.
109. Alberts, B., et al., *Molecular Biology of the Cell*. 4th edition ed. 2002, New York: Garland Science.

110. Graham, J.M., *Isolation of lysosomes from tissues and cells by differential and density gradient centrifugation.* Curr Protoc Cell Biol, 2001. **7**(1): p. 3.6.1-A.3D.8.
111. Leighton, F., et al., *The large-scale separation of peroxisomes, mitochondria, and lysosomes from the livers of rats injected with triton WR-1339. Improved isolation procedures, automated analysis, biochemical and morphological properties of fractions.* J Cell Biol, 1968. **37**(2): p. 482-513.
112. Ponsford, A.H., et al., *Live imaging of intra-lysosome pH in cell lines and primary neuronal culture using a novel genetically encoded biosensor.* Autophagy, 2021. **17**(6): p. 1500-1518.
113. Abu-Remaileh, M., et al., *Lysosomal metabolomics reveals V-ATPase- and mTOR-dependent regulation of amino acid efflux from lysosomes.* Science, 2017. **358**(6364): p. 807-813.
114. Xiong, J., et al., *Rapid affinity purification of intracellular organelles using a twin strep tag.* J Cell Sci, 2019. **132**(24): p. jcs.235390.
115. Walker, M.W. and E. Lloyd-Evans, *A rapid method for the preparation of ultrapure, functional lysosomes using functionalized superparamagnetic iron oxide nanoparticles.* Methods Cell Biol, 2015. **126**: p. 21-43.
116. Wendeler, M. and K. Sandhoff, *Hexosaminidase assays.* Glycoconj J, 2009. **26**(8): p. 945-52.
117. Singh, J., et al., *Systematic Comparison of Strategies for the Enrichment of Lysosomes by Data Independent Acquisition.* J Proteome Res, 2020. **19**(1): p. 371-381.
118. Tharkeshwar, A.K., D. Demedts, and W. Annaert, *Superparamagnetic Nanoparticles for Lysosome Isolation to Identify Spatial Alterations in Lysosomal Protein and Lipid Composition.* STAR Protoc, 2020. **1**(3): p. 100122.
119. Arora, D., et al., *Updates on the study of lysosomal protein dynamics: possibilities for the clinic.* Expert Rev Proteomics, 2023. **20**(1-3): p. 47-55.
120. Borràs, E. and E. Sabidó, *What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry.* Proteomics, 2017. **17**(17-18): p. 201700180.
121. Varela, M.A. and A. Schmidt, *The Emerging Potential of Advanced Targeted Mass Spectrometry to Become a Routine Tool for Protein Quantification in Biomedical Research.* Chimia, 2022. **76**(1-2): p. 81-89.
122. Yost, R.A. and C.G. Enke, *Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation.* Anal Chem, 1979. **51**(12): p. 1251-64.
123. Bauer, M., et al., *Evaluation of data-dependent and -independent mass spectrometric workflows for sensitive quantification of proteins and phosphorylation sites.* J Proteome Res, 2014. **13**(12): p. 5973-88.
124. Picotti, P., et al., *Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics.* Cell, 2009. **138**(4): p. 795-806.
125. Shi, T., et al., *Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum.* Proc Natl Acad Sci U S A, 2012. **109**(38): p. 15395-400.
126. Peterson, A.C., et al., *Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics.* Mol Cell Proteomics, 2012. **11**(11): p. 1475-88.
127. Schilling, B., et al., *Multiplexed, Scheduled, High-Resolution Parallel Reaction Monitoring on a Full Scan QqTOF Instrument with Integrated Data-Dependent and Targeted Mass Spectrometric Workflows.* Anal Chem, 2015. **87**(20): p. 10222-9.
128. Gallien, S., et al., *Selectivity of LC-MS/MS analysis: implication for proteomics experiments.* J Proteomics, 2013. **81**: p. 148-58.
129. Ronsein, G.E., et al., *Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics.* J Proteomics, 2015. **113**: p. 388-99.
130. Schiffmann, C., et al., *Comparison of targeted peptide quantification assays for reductive dehalogenases by selective reaction monitoring (SRM) and precursor reaction monitoring (PRM).* Anal Bioanal Chem, 2014. **406**(1): p. 283-91.
131. Lange, V., et al., *Selected reaction monitoring for quantitative proteomics: a tutorial.* Mol Syst Biol, 2008. **4**: p. 222.
132. Stahl-Zeng, J., et al., *High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites.* Mol Cell Proteomics, 2007. **6**(10): p. 1809-17.
133. Liebler, D.C. and L.J. Zimmerman, *Targeted quantitation of proteins by mass spectrometry.* Biochemistry, 2013. **52**(22): p. 3797-806.

134. Bereman, M.S., et al., *The development of selected reaction monitoring methods for targeted proteomics via empirical refinement.* Proteomics, 2012. **12**(8): p. 1134-41.
135. Picotti, P. and R. Aebersold, *Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions.* Nat Methods, 2012. **9**(6): p. 555-66.
136. Vidova, V. and Z. Spacil, *A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition.* Anal Chim Acta, 2017. **964**: p. 7-23.
137. Kulyyassov, A., M. Fresnais, and R. Longuespée, *Targeted liquid chromatography-tandem mass spectrometry analysis of proteins: Basic principles, applications, and perspectives.* Proteomics, 2021. **21**(23-24): p. e2100153.
138. Kuster, B., et al., *Scoring proteomes with proteotypic peptide probes.* Nat Rev Mol Cell Biol, 2005. **6**(7): p. 577-83.
139. Mallick, P., et al., *Computational prediction of proteotypic peptides for quantitative proteomics.* Nat Biotechnol, 2007. **25**(1): p. 125-31.
140. Kirkpatrick, D.S., S.A. Gerber, and S.P. Gygi, *The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications.* Methods, 2005. **35**(3): p. 265-73.
141. Zolg, D.P., et al., *Building ProteomeTools based on a complete synthetic human proteome.* Nat Methods, 2017. **14**(3): p. 259-262.
142. Sharma, V., et al., *Panorama Public: A Public Repository for Quantitative Data Sets Processed in Skyline.* Mol Cell Proteomics, 2018. **17**(6): p. 1239-1244.
143. Bhowmick, P., Y. Mohammed, and C.H. Borchers, *MRMAssayDB: an integrated resource for validated targeted proteomics assays.* Bioinformatics, 2018. **34**(20): p. 3566-3571.
144. Bhowmick, P., et al., *An Update on MRMAssayDB: A Comprehensive Resource for Targeted Proteomics Assays in the Community.* J Proteome Res, 2021. **20**(4): p. 2105-2115.
145. Zauber, H., M. Kirchner, and M. Selbach, *Picky: a simple online PRM and SRM method designer for targeted proteomics.* Nat Methods, 2018. **15**(3): p. 156-157.
146. Mohammed, Y., et al., *PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments.* J Proteomics, 2014. **106**: p. 151-61.
147. Picotti, P., et al., *High-throughput generation of selected reaction-monitoring assays for proteins and proteomes.* Nat Methods, 2010. **7**(1): p. 43-6.
148. Escher, C., et al., *Using iRT, a normalized retention time for more targeted measurement of peptides.* Proteomics, 2012. **12**(8): p. 1111-21.
149. Unwin, R.D., J.R. Griffiths, and A.D. Whetton, *A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: multiple reaction monitoring-initiated detection and sequencing (MIDAS).* Nat Protoc, 2009. **4**(6): p. 870-7.
150. Maclean, B., et al., *Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry.* Anal Chem, 2010. **82**(24): p. 10116-24.
151. Pino, L.K., et al., *The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics.* Mass Spectrom Rev, 2020. **39**(3): p. 229-244.
152. Zhu, S., et al., *Targeted Multiple Reaction Monitoring Analysis of CSF Identifies UCHL1 and GPNMB as Candidate Biomarkers for ALS.* J Mol Neurosci, 2019. **69**(4): p. 643-657.
153. Heywood, W.E., et al., *Identification of novel CSF biomarkers for neurodegeneration and their validation by a high-throughput multiplexed targeted proteomic assay.* Mol Neurodegener, 2015. **10**: p. 64.
154. Sjödin, S., et al., *Endo-lysosomal proteins and ubiquitin CSF concentrations in Alzheimer's and Parkinson's disease.* Alzheimers Res Ther, 2019. **11**(1): p. 82.
155. Lowe, A.J., et al., *Cerebrospinal fluid endo-lysosomal proteins as potential biomarkers for Huntington's disease.* PLoS One, 2020. **15**(8): p. e0233820.
156. Zhang, T., et al., *A rapid and non-invasive proteomic analysis using DBS and buccal swab for multiplexed second-tier screening of Pompe disease and Mucopolysaccharidosis type I.* Mol Genet Metab, 2022. **136**(4): p. 296-305.
157. Heywood, W.E., et al., *Proteomic Discovery and Development of a Multiplexed Targeted MRM-LC-MS/MS Assay for Urine Biomarkers of Extracellular Matrix Disruption in Mucopolysaccharidoses I, II, and VI.* Anal Chem, 2015. **87**(24): p. 12238-44.
158. Ohira, M., T. Okuyama, and R. Mashima, *Quantification of 11 enzyme activities of lysosomal storage disorders using liquid chromatography-tandem mass spectrometry.* Mol Genet Metab Rep, 2018. **17**: p. 9-15.

159. Liu, K., et al., *Co-Analysis of Serum and Urine Differentially Expressed Proteins in Mucopolysaccharidosis Type I.* J Proteome Res, 2024. **23**(2): p. 718-727.
160. Stoka, V., V. Turk, and B. Turk, *Lysosomal cathepsins and their regulation in aging and neurodegeneration.* Ageing Res Rev, 2016. **32**: p. 22-37.
161. Terasawa, K., et al., *Lysosome-associated membrane proteins-1 and -2 (LAMP-1 and LAMP-2) assemble via distinct modes.* Biochem Biophys Res Commun, 2016. **479**(3): p. 489-495.
162. Gold, L., et al., *Aptamer-based multiplexed proteomic technology for biomarker discovery.* PLoS One, 2010. **5**(12): p. e15004.
163. Inc., S., *SOMAscan Proteomic Assay - Technical White Paper*. 2016. p. 1-14.
164. Assarsson, E., et al., *Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability.* PLoS One, 2014. **9**(4): p. e95192.
165. AB, O.P., *White Paper - PEA – a high-multiplex immunoassay technology with qPCR or NGS readout*. 2020. p. 1-4.
166. AB, O.P. 2024 [cited 2024 08.01.2024]; Available from: https://olink.com.
167. Inc., S. [cited 2024 08.01.2024]; Available from: https://somalogic.com/.
168. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present.* Anal Bioanal Chem, 2012. **404**(4): p. 939-65.
169. Huang, T., et al., *Combining Precursor and Fragment Information for Improved Detection of Differential Abundance in Data Independent Acquisition.* Mol Cell Proteomics, 2020. **19**(2): p. 421-430.
170. Xie, F., et al., *Liquid chromatography-mass spectrometry-based quantitative proteomics.* J Biol Chem, 2011. **286**(29): p. 25443-9.
171. Panuwet, P., et al., *Biological Matrix Effects in Quantitative Tandem Mass Spectrometry-Based Analytical Methods: Advancing Biomonitoring.* Crit Rev Anal Chem, 2016. **46**(2): p. 93-105.
172. Arnold, S.L., F. Stevison, and N. Isoherranen, *Impact of Sample Matrix on Accuracy of Peptide Quantification: Assessment of Calibrator and Internal Standard Selection and Method Validation.* Anal Chem, 2016. **88**(1): p. 746-53.
173. Bondarenko, P.V., D. Chelius, and T.A. Shaler, *Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry.* Anal Chem, 2002. **74**(18): p. 4741-9.
174. Pappireddi, N., L. Martin, and M. Wühr, *A Review on Quantitative Multiplexed Proteomics.* Chembiochem, 2019. **20**(10): p. 1210-1224.
175. Washburn, M.P., D. Wolters, and J.R. Yates, 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotechnol, 2001. **19**(3): p. 242-7.
176. Schwanhäusser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-42.
177. Zhang, R., et al., *Fractionation of isotopically labeled peptides in quantitative proteomics.* Anal Chem, 2001. **73**(21): p. 5142-9.
178. Ye, X., et al., *18O stable isotope labeling in MS-based proteomics.* Brief Funct Genomic Proteomic, 2009. **8**(2): p. 136-44.
179. Gouw, J.W., J. Krijgsveld, and A.J. Heck, *Quantitative proteomics by metabolic labeling of model organisms.* Mol Cell Proteomics, 2010. **9**(1): p. 11-24.
180. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.
181. Chen, X., et al., *Quantitative proteomics using SILAC: Principles, applications, and developments.* Proteomics, 2015. **15**(18): p. 3175-92.
182. Blagoev, B., et al., *Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics.* Nat Biotechnol, 2004. **22**(9): p. 1139-45.
183. Krüger, M., et al., *SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function.* Cell, 2008. **134**(2): p. 353-64.
184. Soufi, B. and B. Macek, *Stable isotope labeling by amino acids applied to bacterial cell culture.* Methods Mol Biol, 2014. **1188**: p. 9-22.
185. de Godoy, L.M., *SILAC yeast: from labeling to comprehensive proteome quantification.* Methods Mol Biol, 2014. **1156**: p. 81-109.
186. Schütz, W., et al., *Extending SILAC to proteomics of plant cell lines.* Plant Cell, 2011. **23**(5): p. 1701-5.
187. Westman-Brinkmalm, A., et al., *SILAC zebrafish for quantitative analysis of protein turnover and tissue regeneration.* J Proteomics, 2011. **75**(2): p. 425-34.

188. Krijgsveld, J., et al., *Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics.* Nat Biotechnol, 2003. **21**(8): p. 927-31.
189. McClatchy, D.B., et al., *15N metabolic labeling of mammalian tissue with slow protein turnover.* J Proteome Res, 2007. **6**(5): p. 2005-10.
190. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Anal Chem, 2003. **75**(8): p. 1895-904.
191. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.
192. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.
193. Kang, U.B., et al., *Quantitative analysis of mTRAQ-labeled proteome using full MS scans.* J Proteome Res, 2010. **9**(7): p. 3750-8.
194. Boersema, P.J., et al., *Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics.* Nat Protoc, 2009. **4**(4): p. 484-94.
195. Li, J., et al., *TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing.* J Proteome Res, 2021. **20**(5): p. 2964-2972.
196. Pottiez, G., et al., *Comparison of 4-plex to 8-plex iTRAQ quantitative measurements of proteins in human plasma samples.* J Proteome Res, 2012. **11**(7): p. 3774-81.
197. McAlister, G.C., et al., *MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes.* Anal Chem, 2014. **86**(14): p. 7150-8.
198. Udeshi, N.D., et al., *Rapid and deep-scale ubiquitylation profiling for biology and translational research.* Nat Commun, 2020. **11**(1): p. 359.
199. Debets, D.O., et al., *Deep (phospho)proteomics profiling of pre- treatment needle biopsies identifies signatures of treatment resistance in HER2(+) breast cancer.* Cell Rep Med, 2023. **4**(10): p. 101203.
200. Winter, D., R. Pipkorn, and W.D. Lehmann, *Separation of peptide isomers and conformers by ultra performance liquid chromatography.* J Sep Sci, 2009. **32**(8): p. 1111-9.
201. Winter, D., et al., *Minimally permutated peptide analogs as internal standards for relative and absolute quantification of peptides and proteins.* Proteomics, 2010. **10**(7): p. 1510-4.
202. Prasad, B. and J.D. Unadkat, *Comparison of Heavy Labeled (SIL) Peptide versus SILAC Protein Internal Standards for LC-MS/MS Quantification of Hepatic Drug Transporters.* Int J Proteomics, 2014. **2014**: p. 451510.
203. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS.* Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6940-5.
204. Schmidt, C., et al., *Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring.* Anal Chem, 2010. **82**(7): p. 2784-96.
205. Tsai, C.F., et al., *Large-scale determination of absolute phosphorylation stoichiometries in human cells by motif-targeting quantitative proteomics.* Nat Commun, 2015. **6**: p. 6622.
206. Pratt, J.M., et al., *Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes.* Nat Protoc, 2006. **1**(2): p. 1029-43.
207. Dhillon, B.K., et al., *Systems Biology Approaches to Understanding the Human Immune System.* Front Immunol, 2020. **11**: p. 1683.
208. Miyauchi, E., et al., *Identification of blood biomarkers in glioblastoma by SWATH mass spectrometry and quantitative targeted absolute proteomics.* PLoS One, 2018. **13**(3): p. e0193799.
209. Rifai, N., M.A. Gillette, and S.A. Carr, *Protein biomarker discovery and validation: the long and uncertain path to clinical utility.* Nat Biotechnol, 2006. **24**(8): p. 971-83.
210. Persson, A., S. Hober, and M. Uhlén, *A human protein atlas based on antibody proteomics.* Curr Opin Mol Ther, 2006. **8**(3): p. 185-90.
211. Uhlén, M., et al., *A human protein atlas for normal and cancer tissues based on antibody proteomics.* Mol Cell Proteomics, 2005. **4**(12): p. 1920-32.
212. Zeiler, M., et al., *A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines.* Mol Cell Proteomics, 2012. **11**(3): p. O111.009613.
213. Hanke, S., et al., *Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level.* J Proteome Res, 2008. **7**(3): p. 1118-30.

214. Singh, S., et al., *FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides.* J Proteome Res, 2009. **8**(5): p. 2201-10.
215. Brun, V., et al., *Isotope-labeled protein standards: toward absolute quantitative proteomics.* Mol Cell Proteomics, 2007. **6**(12): p. 2139-49.
216. Qasrawi, D.O., E.V. Petrotchenko, and C.H. Borchers, *Amino acid analysis for peptide quantitation using reversed-phase liquid chromatography combined with multiple reaction monitoring mass spectrometry.* Anal Bioanal Chem, 2023. **415**(22): p. 5261-5267.
217. Kettenbach, A.N., J. Rush, and S.A. Gerber, *Absolute quantification of protein and post-translational modification abundance with stable isotope-labeled synthetic peptides.* Nat Protoc, 2011. **6**(2): p. 175-86.
218. Takemori, N., et al., *MEERCAT: Multiplexed Efficient Cell Free Expression of Recombinant QconCATs For Large Scale Absolute Proteome Quantification.* Mol Cell Proteomics, 2017. **16**(12): p. 2169-2183.
219. Lawless, C., et al., *Direct and Absolute Quantification of over 1800 Yeast Proteins via Selected Reaction Monitoring.* Mol Cell Proteomics, 2016. **15**(4): p. 1309-22.
220. Brownridge, P. and R.J. Beynon, *The importance of the digest: proteolysis and absolute quantification in proteomics.* Methods, 2011. **54**(4): p. 351-60.
221. Kito, K., et al., *A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes.* J Proteome Res, 2007. **6**(2): p. 792-800.
222. León, I.R., et al., *Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis.* Mol Cell Proteomics, 2013. **12**(10): p. 2992-3005.
223. Winter, D. and H. Steen, *Optimization of cell lysis and protein digestion protocols for the analysis of HeLa S3 cells by LC-MS/MS.* Proteomics, 2011. **11**(24): p. 4726-30.
224. Shevchenko, A., et al., *In-gel digestion for mass spectrometric characterization of proteins and proteomes.* Nat Protoc, 2006. **1**(6): p. 2856-60.
225. Hughes, C.S., et al., *Ultrasensitive proteome analysis using paramagnetic bead technology.* Mol Syst Biol, 2014. **10**(10): p. 757.
226. Hughes, C.S., et al., *Single-pot, solid-phase-enhanced sample preparation for proteomics experiments.* Nat Protoc, 2019. **14**(1): p. 68-85.
227. Bennike, T.B., et al., *A Cost-Effective High-Throughput Plasma and Serum Proteomics Workflow Enables Mapping of the Molecular Impact of Total Pancreatectomy with Islet Autotransplantation.* J Proteome Res, 2018. **17**(5): p. 1983-1992.
228. Wiśniewski, J.R., et al., *Universal sample preparation method for proteome analysis.* Nat Methods, 2009. **6**(5): p. 359-62.
229. Zougman, A., P.J. Selby, and R.E. Banks, *Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis.* Proteomics, 2014. **14**(9): p. 1006-0.
230. Kulak, N.A., et al., *Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells.* Nat Methods, 2014. **11**(3): p. 319-24.
231. Mosen, P.R., R. Hardt, and D. Winter, *RapiGest precipitation depends on peptide concentration.* Proteomics, 2021. **21**(20): p. e2100129.
232. Yu, Y.Q., et al., *Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins.* Anal Chem, 2003. **75**(21): p. 6023-8.
233. Michalski, A., J. Cox, and M. Mann, *More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS.* J Proteome Res, 2011. **10**(4): p. 1785-93.
234. Egertson, J.D., et al., *Multiplexed peptide analysis using data-independent acquisition and Skyline.* Nat Protoc, 2015. **10**(6): p. 887-903.
235. Zhang, F., et al., *A Comparative Analysis of Data Analysis Tools for Data-Independent Acquisition Mass Spectrometry.* Mol Cell Proteomics, 2023. **22**(9): p. 100623.
236. Bruderer, R., et al., *Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues.* Mol Cell Proteomics, 2015. **14**(5): p. 1400-10.
237. Bekker-Jensen, D.B., et al., *Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries.* Nat Commun, 2020. **11**(1): p. 787.
238. Kockmann, T., et al., *Targeted proteomics coming of age - SRM, PRM and DIA performance evaluated from a core facility perspective.* Proteomics, 2016. **16**(15-16): p. 2183-92.

239. Martins-de-Souza, D., V.M. Faça, and F.C. Gozzo, *DIA is not a new mass spectrometry acquisition method.* Proteomics, 2017. **17**(7): p. 00017.
240. Guo, T. and R. Aebersold, *Recent advances of data-independent acquisition mass spectrometry-based proteomics.* Proteomics, 2023. **23**(7-8): p. e2200011.
241. Ponnaiyan, S., et al., *Comprehensive draft of the mouse embryonic fibroblast lysosomal proteome by mass spectrometry based proteomics.* Sci Data, 2020. **7**(1): p. 68.
242. Kverneland, A.H., et al., *Differential ultracentrifugation enables deep plasma proteomics through enrichment of extracellular vesicles.* Proteomics, 2023. **23**(7-8): p. e2200039.
243. Karayel, O., et al., *DIA-based systems biology approach unveils E3 ubiquitin ligase-dependent responses to a metabolic shift.* Proc Natl Acad Sci U S A, 2020. **117**(51): p. 32806-32815.
244. Weke, K., et al., *DIA-MS proteome analysis of formalin-fixed paraffin-embedded glioblastoma tissues.* Anal Chim Acta, 2022. **1204**: p. 339695.
245. Li, K.W., et al., *Recent Developments in Data Independent Acquisition (DIA) Mass Spectrometry: Application of Quantitative Analysis of the Brain Proteome.* Front Mol Neurosci, 2020. **13**: p. 564446.
246. Collins, B.C., et al., *Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry.* Nat Commun, 2017. **8**(1): p. 291.
247. Liu, Y., et al., *Quantitative variability of 342 plasma proteins in a human twin population.* Mol Syst Biol, 2015. **11**(1): p. 786.
248. Lou, R., et al., *Benchmarking commonly used software suites and analysis workflows for DIA proteomics and phosphoproteomics.* Nat Commun, 2023. **14**(1): p. 94.
249. Schmidlin, T., et al., *Assessment of SRM, MRM(3) , and DIA for the targeted analysis of phosphorylation dynamics in non-small cell lung cancer.* Proteomics, 2016. **16**(15-16): p. 2193-205.
250. Liu, Y., et al., *Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS.* Proteomics, 2013. **13**(8): p. 1247-56.
251. Dillen, L., et al., *Comparison of triple quadrupole and high-resolution TOF-MS for quantification of peptides.* Bioanalysis, 2012. **4**(5): p. 565-79.
252. Li, C., et al., *Towards Higher Sensitivity of Mass Spectrometry: A Perspective From the Mass Analyzers.* Front Chem, 2021. **9**: p. 813359.
253. Reubsaet, L., M.J. Sweredoski, and A. Moradian, *Data-Independent Acquisition for the Orbitrap Q Exactive HF: A Tutorial.* J Proteome Res, 2019. **18**(3): p. 803-813.
254. Pino, L.K., et al., *Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries.* Mol Cell Proteomics, 2020. **19**(7): p. 1088-1103.
255. Gotti, C., et al., *Extensive and Accurate Benchmarking of DIA Acquisition Methods and Software Tools Using a Complex Proteomic Standard.* J Proteome Res, 2021. **20**(10): p. 4801-4814.
256. Zubarev, R.A. and A. Makarov, *Orbitrap mass spectrometry.* Anal Chem, 2013. **85**(11): p. 5288-96.
257. Johnson, D.E., et al., *The position of lysosomes within the cell determines their luminal pH.* J Cell Biol, 2016. **212**(6): p. 677-92.
258. Ballabio, A. and J.S. Bonifacino, *Lysosomes as dynamic regulators of cell and organismal homeostasis.* Nat Rev Mol Cell Biol, 2020. **21**(2): p. 101-118.
259. Yu, Y., et al., *Organelle proteomic profiling reveals lysosomal heterogeneity in association with longevity.* Elife, 2024. **13**: p. 85214.
260. Nilsson, E., R. Ghassemifar, and U.T. Brunk, *Lysosomal heterogeneity between and within cells with respect to resistance against oxidative stress.* Histochem J, 1997. **29**(11-12): p. 857-65.
261. Goldstone, A. and H. Koenig, *Synthesis and turnover of lysosomal glycoproteins. Relation to the molecular heterogeneity of the lysosomal enzymes.* FEBS Lett, 1974. **39**(2): p. 176-81.
262. Pertoft, H., B. Wärmegård, and M. Höök, *Heterogeneity of lysosomes originating from rat liver parenchymal cells. Metabolic relationship of subpopulations separated by density-gradient centrifugation.* Biochem J, 1978. **174**(1): p. 309-17.
263. Laqtom, N.N., et al., *CLN3 is required for the clearance of glycerophosphodiesters from lysosomes.* Nature, 2022. **609**(7929): p. 1005-1011.
264. Markmann, S., et al., *Quantitative Proteome Analysis of Mouse Liver Lysosomes Provides Evidence for Mannose 6-phosphate-independent Targeting Mechanisms of Acid Hydrolases in Mucolipidosis II.* Mol Cell Proteomics, 2017. **16**(3): p. 438-450.

265. Zigdon, H., A. Meshcheriakova, and A.H. Futerman, *From sheep to mice to cells: tools for the study of the sphingolipidoses.* Biochim Biophys Acta, 2014. **1841**(8): p. 1189-99.
266. Richter, M., et al., *From Donor to the Lab: A Fascinating Journey of Primary Cell Lines.* Front Cell Dev Biol, 2021. **9**: p. 711381.
267. Harper, J.M., *Primary Cell Culture as a Model System for Evolutionary Molecular Physiology.* Int J Mol Sci, 2024. **25**(14): p. 7905.
268. Zhao, C., *Cell culture: in vitro model system and a promising path to in vivo applications.* J Histotechnol, 2023. **46**(1): p. 1-4.
269. Carr, S.A., et al., *Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach.* Mol Cell Proteomics, 2014. **13**(3): p. 907-17.
270. Sharma, K., et al., *Cell type- and brain region-resolved mouse brain proteome.* Nat Neurosci, 2015. **18**(12): p. 1819-31.
271. Azimifar, S.B., et al., *Cell-type-resolved quantitative proteomics of murine liver.* Cell Metab, 2014. **20**(6): p. 1076-87.
272. Sleat, D.E., et al., *Extending the mannose 6-phosphate glycoproteome by high resolution/accuracy mass spectrometry analysis of control and acid phosphatase 5-deficient mice.* Mol Cell Proteomics, 2013. **12**(7): p. 1806-17.
273. Huttlin, E.L., et al., *A tissue-specific atlas of mouse protein phosphorylation and expression.* Cell, 2010. **143**(7): p. 1174-89.
274. Wilkins, M.R., et al., *Protein identification and analysis tools in the ExPASy server.* Methods Mol Biol, 1999. **112**: p. 531-52.
275. Austin, B.P., S. Nallamsetty, and D.S. Waugh, *Hexahistidine-tagged maltose-binding protein as a fusion partner for the production of soluble recombinant proteins in Escherichia coli.* Methods Mol Biol, 2009. **498**: p. 157-72.
276. Parikh, M.R., et al., *Directed evolution of RuBisCO hypermorphs through genetic selection in engineered E.coli.* Protein Eng Des Sel, 2006. **19**(3): p. 113-9.
277. Mosen, P., et al., *Targeted Quantification of the Lysosomal Proteome in Complex Samples.* Proteomes, 2021. **9**(1): p. 9010004.
278. Wiśniewski, J.R., *Filter-Aided Sample Preparation for Proteome Analysis.* Methods Mol Biol, 2018. **1841**: p. 3-10.
279. Reales-Calderón, J.A., et al., *Candida albicans Modifies the Protein Composition and Size Distribution of THP-1 Macrophage-Derived Extracellular Vesicles.* J Proteome Res, 2017. **16**(1): p. 87-105.
280. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.
281. Wessel, D. and U.I. Flügge, *A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids.* Anal Biochem, 1984. **138**(1): p. 141-3.
282. Müller, T. and D. Winter, *Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents.* Mol Cell Proteomics, 2017. **16**(7): p. 1173-1187.
283. Rappsilber, J., M. Mann, and Y. Ishihama, *Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.* Nat Protoc, 2007. **2**(8): p. 1896-906.
284. MacLean, B., et al., *Skyline: an open source document editor for creating and analyzing targeted proteomics experiments.* Bioinformatics, 2010. **26**(7): p. 966-8.
285. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.
286. Chambers, M.C., et al., *A cross-platform toolkit for mass spectrometry and proteomics.* Nat Biotechnol, 2012. **30**(10): p. 918-20.
287. Käll, L., et al., *Semi-supervised learning for peptide identification from shotgun proteomics datasets.* Nat Methods, 2007. **4**(11): p. 923-5.
288. Bereman, M.S., et al., *An Automated Pipeline to Monitor System Performance in Liquid Chromatography-Tandem Mass Spectrometry Proteomic Experiments.* J Proteome Res, 2016. **15**(12): p. 4763-4769.
289. Reiter, L., et al., *mProphet: automated data processing and statistical validation for large-scale SRM experiments.* Nat Methods, 2011. **8**(5): p. 430-5.

290. R Core Development Team, *R: A language and environment for statistical computing*. 2010, R Foundation for Statistical Computing: Vienna, Austria.
291. Berg, S., et al., *ilastik: interactive machine learning for (bio)image analysis.* Nat Methods, 2019. **16**(12): p. 1226-1232.
292. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
293. Westermann, L.M., et al., *Imbalanced cellular metabolism compromises cartilage homeostasis and joint function in a mouse model of mucolipidosis type III gamma.* Dis Model Mech, 2020. **13**(11): p. 046425.
294. Di Lorenzo, G., et al., *Lysosomal Proteome and Secretome Analysis Identifies Missorted Enzymes and Their Nondegraded Substrates in Mucolipidosis III Mouse Cells.* Mol Cell Proteomics, 2018. **17**(8): p. 1612-1626.
295. Jungnickel, K.E.J., et al., *MFSD1 with its accessory subunit GLMP functions as a general dipeptide uniporter in lysosomes.* Nat Cell Biol, 2024. **26**(7): p. 1047-1061.
296. Schwake, M., B. Schröder, and P. Saftig, *Lysosomal membrane proteins and their central role in physiology.* Traffic, 2013. **14**(7): p. 739-48.
297. Wilke, S., J. Krausze, and K. Büssow, *Crystal structure of the conserved domain of the DC lysosomal associated membrane protein: implications for the lysosomal glycocalyx.* BMC Biol, 2012. **10**: p. 62.
298. Harrison, M.A. and S.P. Muench, *The Vacuolar ATPase - A Nano-scale Motor That Drives Cell Biology.* Subcell Biochem, 2018. **87**: p. 409-459.
299. Medina, D.L., et al., *Lysosomal calcium signalling regulates autophagy through calcineurin and TFEB.* Nat Cell Biol, 2015. **17**(3): p. 288-99.
300. Lee, C., et al., *The lysosomal potassium channel TMEM175 adopts a novel tetrameric architecture.* Nature, 2017. **547**(7664): p. 472-475.
301. Leisle, L., et al., *ClC-7 is a slowly voltage-gated 2Cl(-)/1H(+)-exchanger and requires Ostm1 for transport activity.* Embo j, 2011. **30**(11): p. 2140-52.
302. Kishimoto, Y., M. Hiraiwa, and J.S. O'Brien, *Saposins: structure, function, distribution, and molecular genetics.* J Lipid Res, 1992. **33**(9): p. 1255-67.
303. Hill, C.H., et al., *The mechanism of glycosphingolipid degradation revealed by a GALC-SapA complex structure.* Nat Commun, 2018. **9**(1): p. 151.
304. Braulke, T. and J.S. Bonifacino, *Sorting of lysosomal proteins.* Biochim Biophys Acta, 2009. **1793**(4): p. 605-14.
305. Chapel, A., et al., *An extended proteome map of the lysosomal membrane reveals novel potential transporters.* Mol Cell Proteomics, 2013. **12**(6): p. 1572-88.
306. Hiraiwa, M., et al., *Lysosomal proteolysis of prosaposin, the precursor of saposins (sphingolipid activator proteins): its mechanism and inhibition by ganglioside.* Arch Biochem Biophys, 1997. **341**(1): p. 17-24.
307. Ebner, M., et al., *Nutrient-regulated control of lysosome function by signaling lipid conversion.* Cell, 2023. **186**(24): p. 5328-5346.e26.
308. Pizzagalli, M.D., A. Bensimon, and G. Superti-Furga, *A guide to plasma membrane solute carrier proteins.* Febs j, 2021. **288**(9): p. 2784-2835.
309. Chou, J.Y., H. Sik Jun, and B.C. Mansfield, *The SLC37 family of phosphate-linked sugar phosphate antiporters.* Mol Aspects Med, 2013. **34**(2-3): p. 601-11.
310. Cappello, A.R., et al., *The Physiopathological Role of the Exchangers Belonging to the SLC37 Family.* Front Chem, 2018. **6**: p. 122.
311. Rebsamen, M., et al., *SLC38A9 is a component of the lysosomal amino acid sensing machinery that controls mTORC1.* Nature, 2015. **519**(7544): p. 477-81.
312. Yoshida, A., et al., *SLC36A1-mTORC1 signaling drives acquired resistance to CDK4/6 inhibitors.* Sci Adv, 2019. **5**(9): p. eaax6352.
313. Verdon, Q., et al., *SNAT7 is the primary lysosomal glutamine exporter required for extracellular protein-dependent growth of cancer cells.* Proc Natl Acad Sci U S A, 2017. **114**(18): p. E3602-e3611.
314. Lange, P.F., et al., *ClC-7 requires Ostm1 as a beta-subunit to support bone resorption and lysosomal function.* Nature, 2006. **440**(7081): p. 220-3.
315. Massa López, D., et al., *The lysosomal transporter MFSD1 is essential for liver homeostasis and critically depends on its accessory subunit GLMP.* Elife, 2019. **8**: p. 50025.

316. Baba, K., et al., *Different localization of lysosomal-associated membrane protein 1 (LAMP1) in mammalian cultured cell lines.* Histochem Cell Biol, 2020. **153**(4): p. 199-213.
317. Wang, F., R. Gómez-Sintes, and P. Boya, *Lysosomal membrane permeabilization and cell death.* Traffic, 2018. **19**(12): p. 918-931.
318. Müller, S., J. Dennemärker, and T. Reinheckel, *Specific functions of lysosomal proteases in endocytic and autophagic pathways.* Biochim Biophys Acta, 2012. **1824**(1): p. 34-43.
319. Gallwitz, L., et al., *Cellular depletion of major cathepsin proteases reveals their concerted activities for lysosomal proteolysis.* Cell Mol Life Sci, 2024. **81**(1): p. 227.
320. Xie, Z., et al., *Cathepsin B in programmed cell death machinery: mechanisms of execution and regulatory pathways.* Cell Death Dis, 2023. **14**(4): p. 255.
321. Ni, J., et al., *Extralysosomal cathepsin B in central nervous system: Mechanisms and therapeutic implications.* Brain Pathol, 2022. **32**(5): p. e13071.
322. Kowalewski, B., et al., *Molecular characterization of arylsulfatase G: expression, processing, glycosylation, transport, and activity.* J Biol Chem, 2014. **289**(40): p. 27992-8005.
323. Frese, M.A., S. Schulz, and T. Dierks, *Arylsulfatase G, a novel lysosomal sulfatase.* J Biol Chem, 2008. **283**(17): p. 11388-95.
324. Lepperdinger, G., J. Müllegger, and G. Kreil, *Hyal2--less active, but more versatile?* Matrix Biol, 2001. **20**(8): p. 509-14.
325. Rai, S.K., et al., *Candidate tumor suppressor HYAL2 is a glycosylphosphatidylinositol (GPI)-anchored cell-surface receptor for jaagsiekte sheep retrovirus, the envelope protein of which mediates oncogenic transformation.* Proc Natl Acad Sci U S A, 2001. **98**(8): p. 4443-8.
326. Chow, G., C.B. Knudson, and W. Knudson, *Expression and cellular localization of human hyaluronidase-2 in articular chondrocytes and cultured cell lines.* Osteoarthritis Cartilage, 2006. **14**(9): p. 849-58.
327. Miller, A.D., et al., *Letter to the Editor: Hyal2, where are you?* Osteoarthritis and Cartilage, 2006. **14**(12): p. 1315-1317.
328. Liu, Y., et al., *Multi-omic measurements of heterogeneity in HeLa cells across laboratories.* Nat Biotechnol, 2019. **37**(3): p. 314-322.
329. Bonini, S. and D. Winter, *Two-Step Enrichment Facilitates Background Reduction for Proteomic Analysis of Lysosomes.* J Proteome Res, 2024. **23**(8): p. 3393-3403.
330. de Juan, A. and B. Lavin Plaza, *Isolation and Culturing of Primary Mouse and Human Macrophages.* Methods Mol Biol, 2022. **2419**: p. 113-124.
331. Fuentes-Mateos, R., E. Santos, and A. Fernández-Medarde, *Optimized Protocol for Isolation and Culture of Murine Neonatal Primary Lung Fibroblasts.* Methods Protoc, 2023. **6**(1): p. 6010014.
332. Chevalier, C., et al., *Primary mouse osteoblast and osteoclast culturing and analysis.* STAR Protoc, 2021. **2**(2): p. 100452.
333. Heybrock, S., et al., *Lysosomal integral membrane protein-2 (LIMP-2/SCARB2) is involved in lysosomal cholesterol export.* Nat Commun, 2019. **10**(1): p. 3521.
334. Pan, C.J., et al., *SLC37A1 and SLC37A2 are phosphate-linked, glucose-6-phosphate antiporters.* PLoS One, 2011. **6**(9): p. e23157.
335. Van den Bossche, F., et al., *Residence of the Nucleotide Sugar Transporter Family Members SLC35F1 and SLC35F6 in the Endosomal/Lysosomal Pathway.* Int J Mol Sci, 2024. **25**(12): p. 25126718.
336. Wyant, G.A., et al., *NUFIP1 is a ribosome receptor for starvation-induced ribophagy.* Science, 2018. **360**(6390): p. 751-758.
337. Meng, D., et al., *SNAT7 regulates mTORC1 via macropinocytosis.* Proc Natl Acad Sci U S A, 2022. **119**(20): p. e2123261119.
338. Koutsifeli, P., et al., *Glycogen-autophagy: Molecular machinery and cellular mechanisms of glycophagy.* J Biol Chem, 2022. **298**(7): p. 102093.
339. Romero, R., et al., *Mechanism of glucocerebrosidase activation and dysfunction in Gaucher disease unraveled by molecular dynamics and deep learning.* Proc Natl Acad Sci U S A, 2019. **116**(11): p. 5086-5095.
340. Turk, B., D. Turk, and V. Turk, *Lysosomal cysteine proteases: more than scavengers.* Biochim Biophys Acta, 2000. **1477**(1-2): p. 98-111.
341. Yan, S., et al., *Transcription of human cathepsin B is mediated by Sp1 and Ets family factors in glioma.* DNA Cell Biol, 2000. **19**(2): p. 79-91.

342. Jean, D., N. Guillaume, and R. Frade, *Characterization of human cathepsin L promoter and identification of binding sites for NF-Y, Sp1 and Sp3 that are essential for its activity.* Biochem J, 2002. **361**(Pt 1): p. 173-84.
343. Motyckova, G., et al., *Linking osteopetrosis and pycnodysostosis: regulation of cathepsin K expression by the microphthalmia transcription factor family.* Proc Natl Acad Sci U S A, 2001. **98**(10): p. 5798-803.
344. Castro-Gomes, T., et al., *Plasma Membrane Repair Is Regulated Extracellularly by Proteases Released from Lysosomes.* PLoS One, 2016. **11**(3): p. e0152583.

8. Supplement

8.1 Supplementary figures

Presented below are the supplementary figures related to Chapter 3.

Figure S1

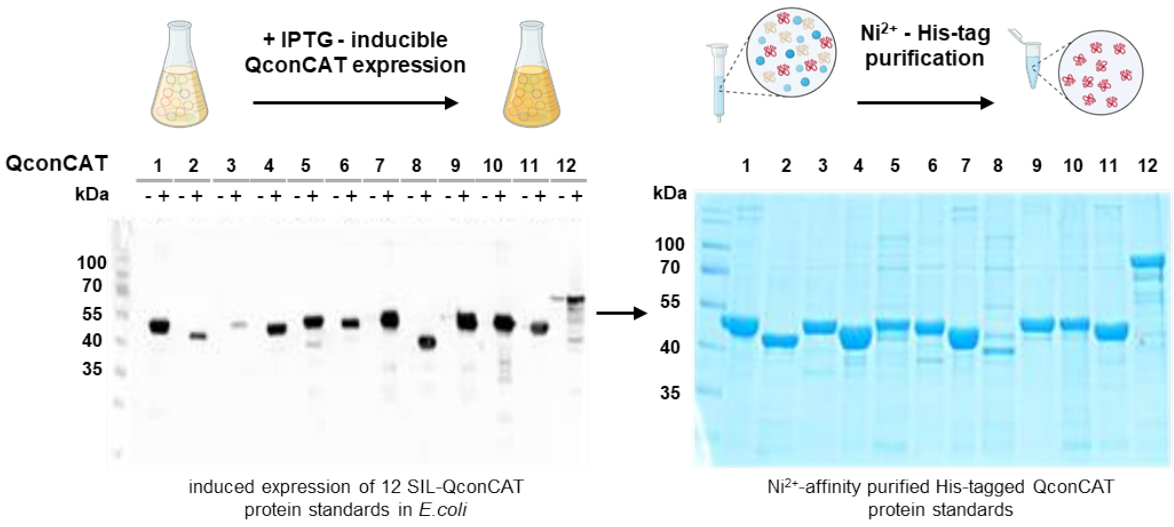


Figure S1: Expression and purification of lysosomal QconCAT protein constructs. Expression of 12 different QconCATs upon IPTG induction (+) was confirmed by Western Blot using a His-tag antibody. (-) shows non-induced controls (left). Heavy stable isotope-labeled (SIL) QconCATs were purified by Ni^{2+}-affinity purification. Abbrev.: IPTG: isopropyl β-D-1-thiogalactopyranoside, SIL: stable isotope-labeled. Figure was generated using BioRender.

Figure S2

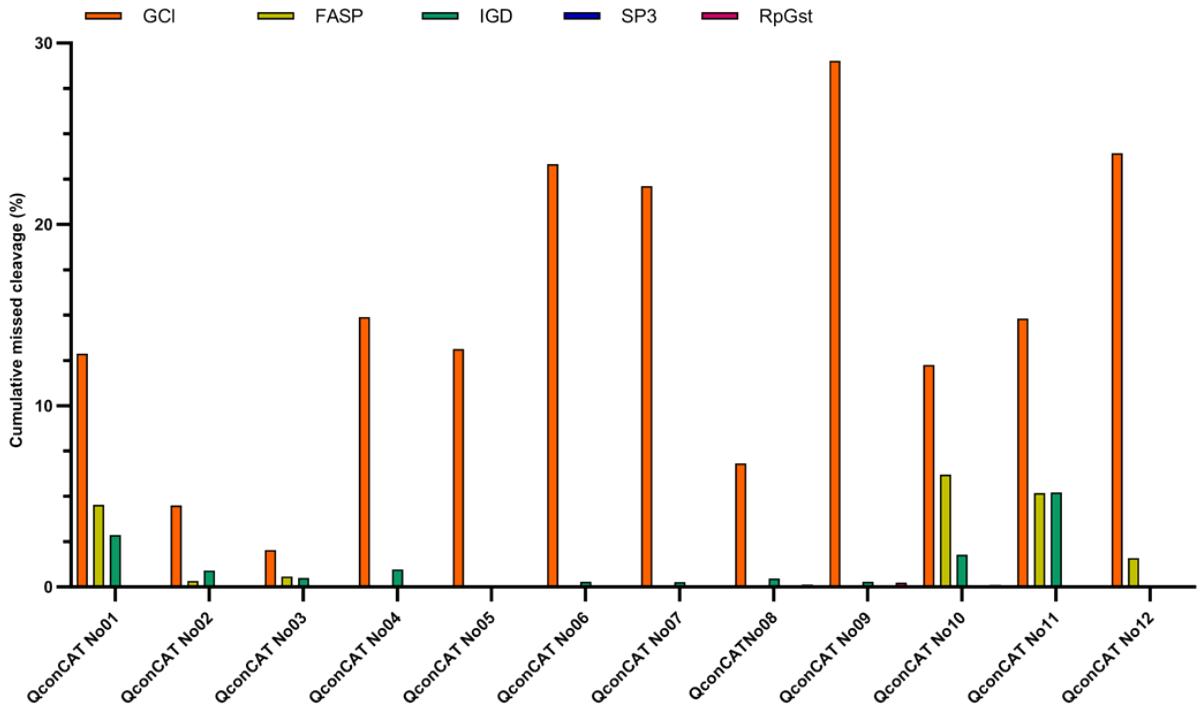
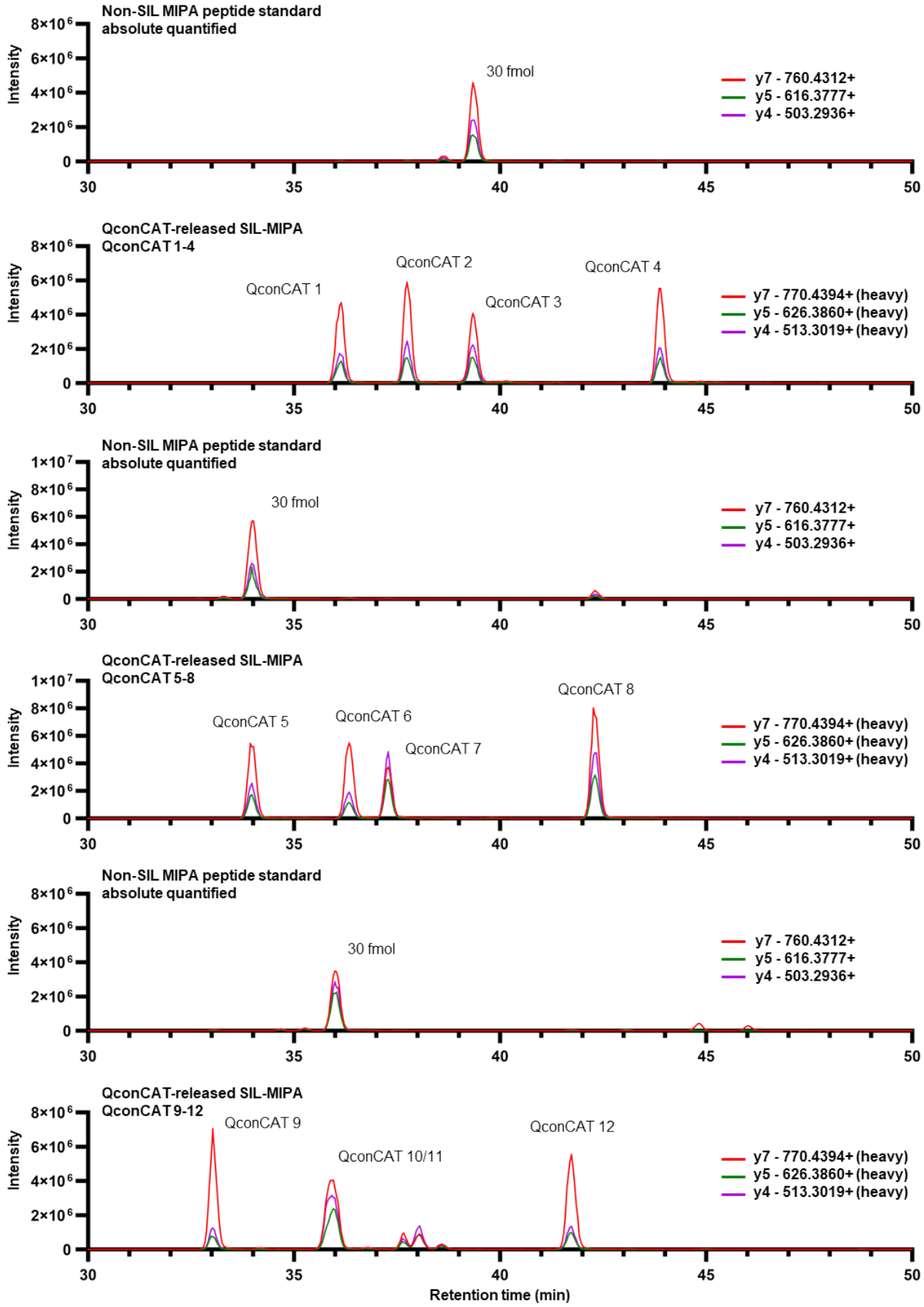
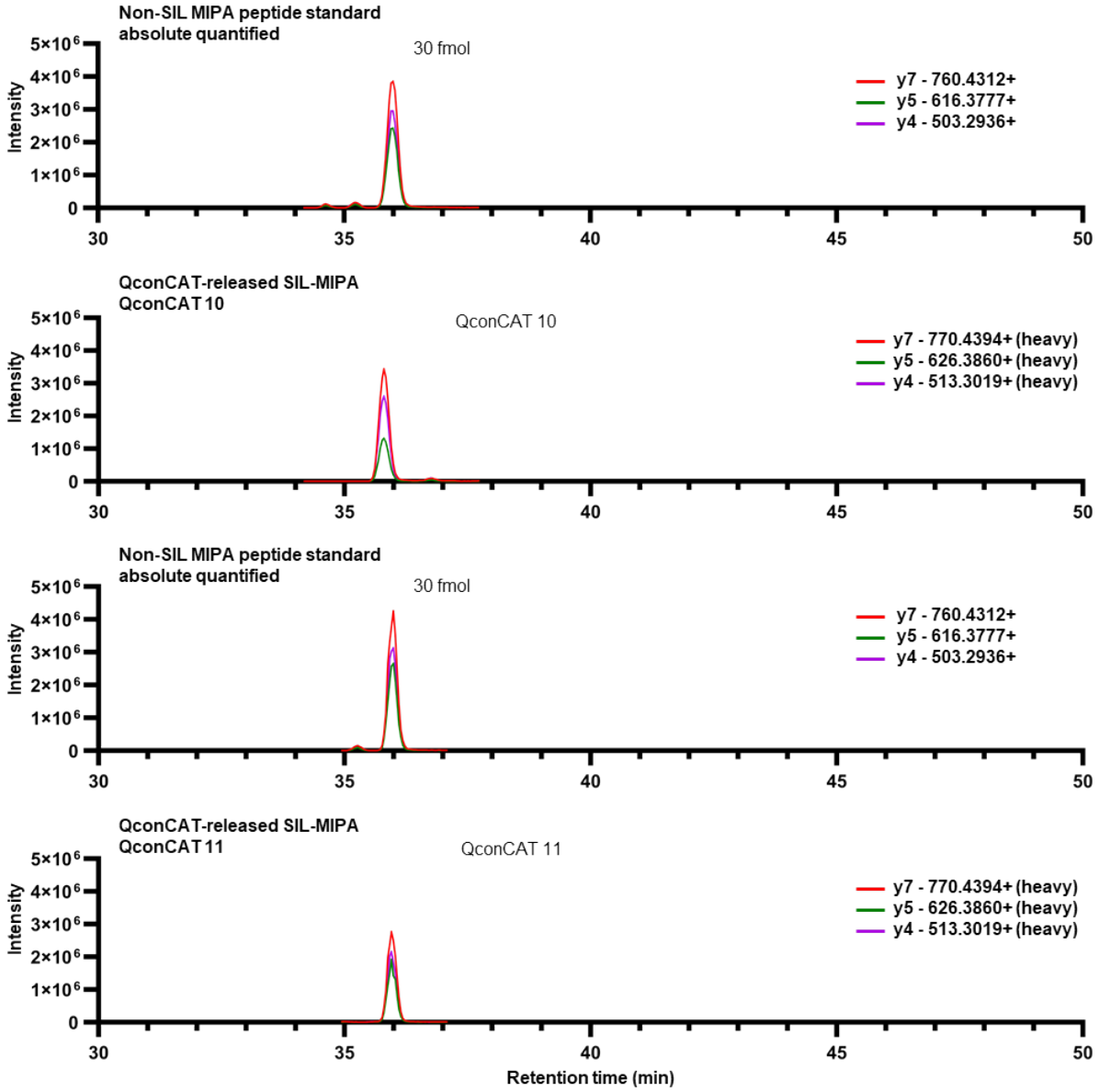


Figure S2: Evaluation and comparison of missed cleavage rates for individual protein standards. Expressed and purified QconCATs were digested using different digestion strategies. Missed cleavage rates reported represent the protein construct's cumulative missed cleavage rate. Abbrev.: IGD: in gel digestion, FASP: Filter Aided Sample Preparation, GCl: guanidinium chlorid, RpGst: RapiGest, SP3: solid-phase-enhanced sample-preparation.

Figure S3


Non-SIL MIPA peptide standard
absolute quantified
30 fmol
y7 - 760.4312+
y5 - 616.3777+
y4 - 503.2936+
QconCAT-released SIL-MIPA
QconCAT 1-4
QconCAT 1
QconCAT 2
QconCAT 3
QconCAT 4
y7 - 770.4394+ (heavy)
y5 - 626.3860+ (heavy)
y4 - 513.3019+ (heavy)
Non-SIL MIPA peptide standard
absolute quantified
30 fmol
y7 - 760.4312+
y5 - 616.3777+
y4 - 503.2936+
QconCAT-released SIL-MIPA
QconCAT 5-8
QconCAT 5
QconCAT 6
QconCAT 7
QconCAT 8
y7 - 770.4394+ (heavy)
y5 - 626.3860+ (heavy)
y4 - 513.3019+ (heavy)
Non-SIL MIPA peptide standard
absolute quantified
30 fmol
y7 - 760.4312+
y5 - 616.3777+
y4 - 503.2936+
QconCAT-released SIL-MIPA
QconCAT 9-12
QconCAT 9
QconCAT 10/11
QconCAT 12
y7 - 770.4394+ (heavy)
y5 - 626.3860+ (heavy)
y4 - 513.3019+ (heavy)
Intensity
Retention time (min)

Continuation Figure S3



Details on the sets of minimally permutated peptide analogs (MIPAs) used in this study - including peptide sequences, QconCAT assignments, isobaric groupings, relative retention times of isobaric peptides, and isotopic precursor m/z values enabling multiplexed absolute re-quantification - are provided in Table S1.

Figure S3: Quantification of 12 stable isotope-labeled QconCAT peptide standards with minimally permutated peptide analogs. Sets of four stable isotope-labeled (SIL) peptide standards (heavy) were quantified against a known amount of non-SIL peptide standard (light). Minimally permutated peptide analogs (MIPAs) representative for QconCAT 10 and 11 co-elute and had to be quantified separately (Continuation Figure S3). Abbrev.: SIL: stable isotope-labeled, MIPA: minimally permutated peptide analog.

Figure S4

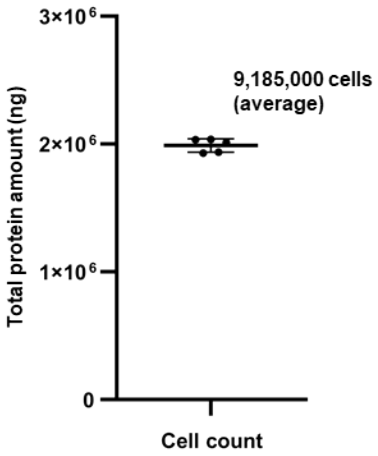


Figure S4: Quantification of protein content in mouse embryonic fibroblast cells. For defined cell numbers lysed the protein content was determined. Shown is the average (n=5), errors bars indicate the standard deviation.

Figure S5

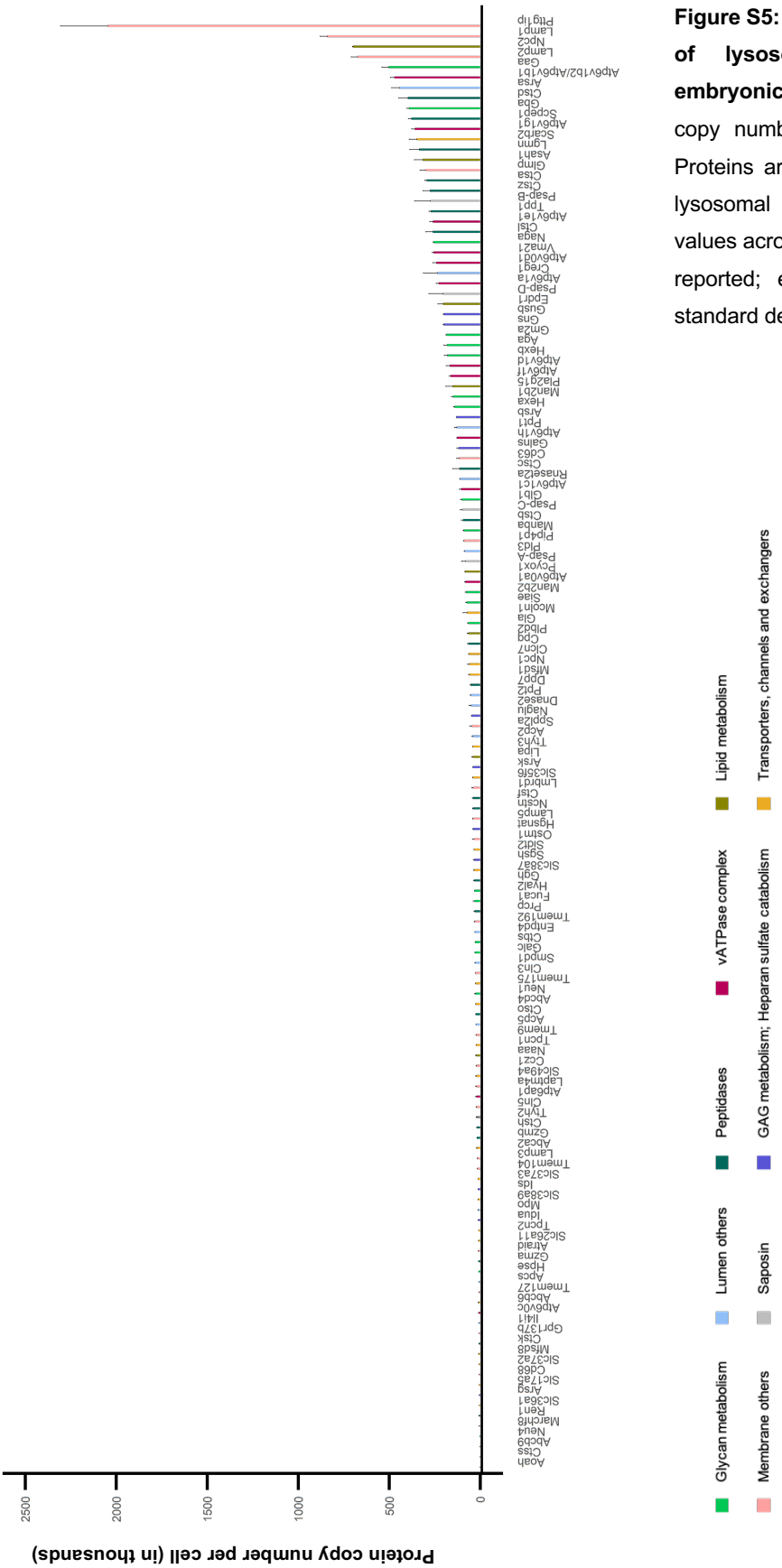
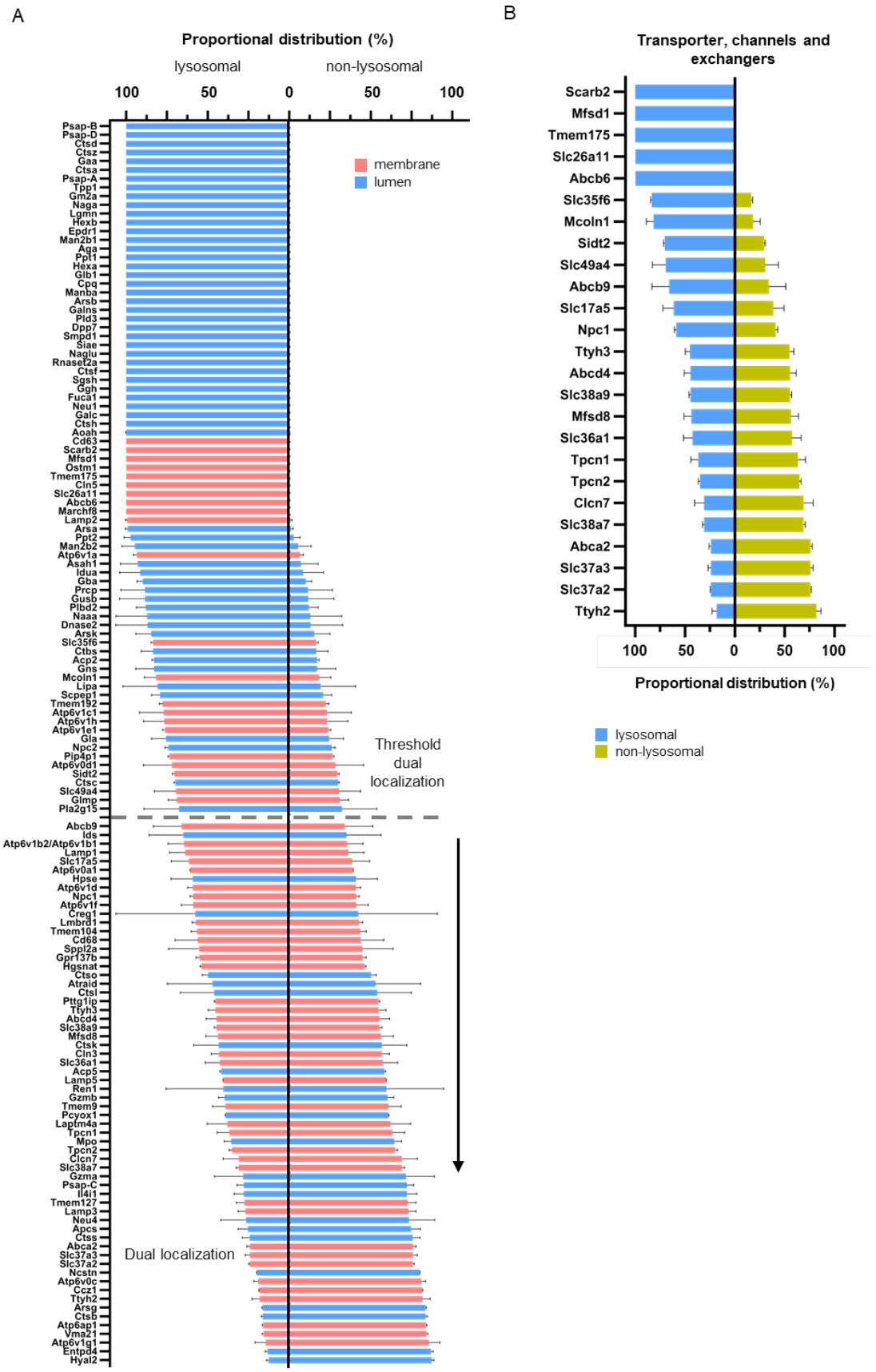


Figure S5: Absolute quantitative analysis of lysosomal proteins in mouse embryonic fibroblasts. Determined protein copy numbers for 141 proteins per cell. Proteins are associated with nine different lysosomal classes/categories. Median values across biological replicates (n=3) are reported; error bars indicate the robust standard deviation (rSD).

Figure S6

Figure S6: Lysosome-specific protein copy number determination from lysosome-enriched fractions. Determined protein copy numbers for 141 proteins per lysosome. Proteins are associated with nine different lysosomal classes/categories. Median values across biological replicates (n=3) are reported; error bars indicate the robust standard deviation (rSD).

Figure S7


A
Proportional distribution (%)
lysosomal
non-lysosomal
100
50
0
50
100
membrane
lumen
Threshold dual localization
Dual localization
B
Transporter, channels and exchangers
Scarb2
Mfsd1
Tmem175
Slc26a11
Abcb6
Slc35f6
Mcoln1
Sidt2
Slc49a4
Abcb9
Slc17a5
Npc1
Ttyh3
Abcd4
Slc38a9
Mfsd8
Slc36a1
Tpcn1
Tpcn2
Clcn7
Slc38a7
Abca2
Slc37a3
Slc37a2
Ttyh2
100
50
0
50
100
Proportional distribution (%)
lysosomal
non-lysosomal

Continuation Figure S7

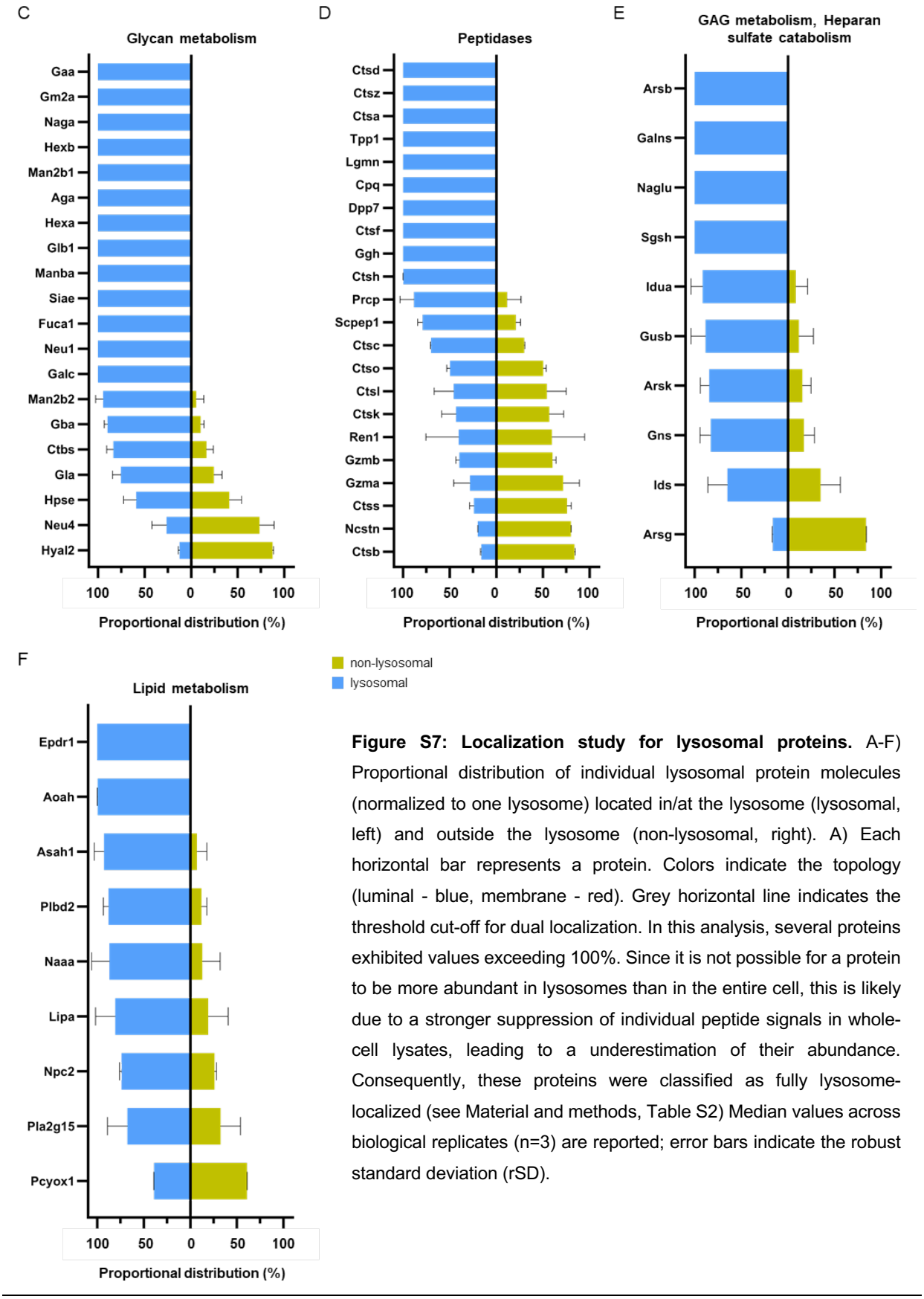
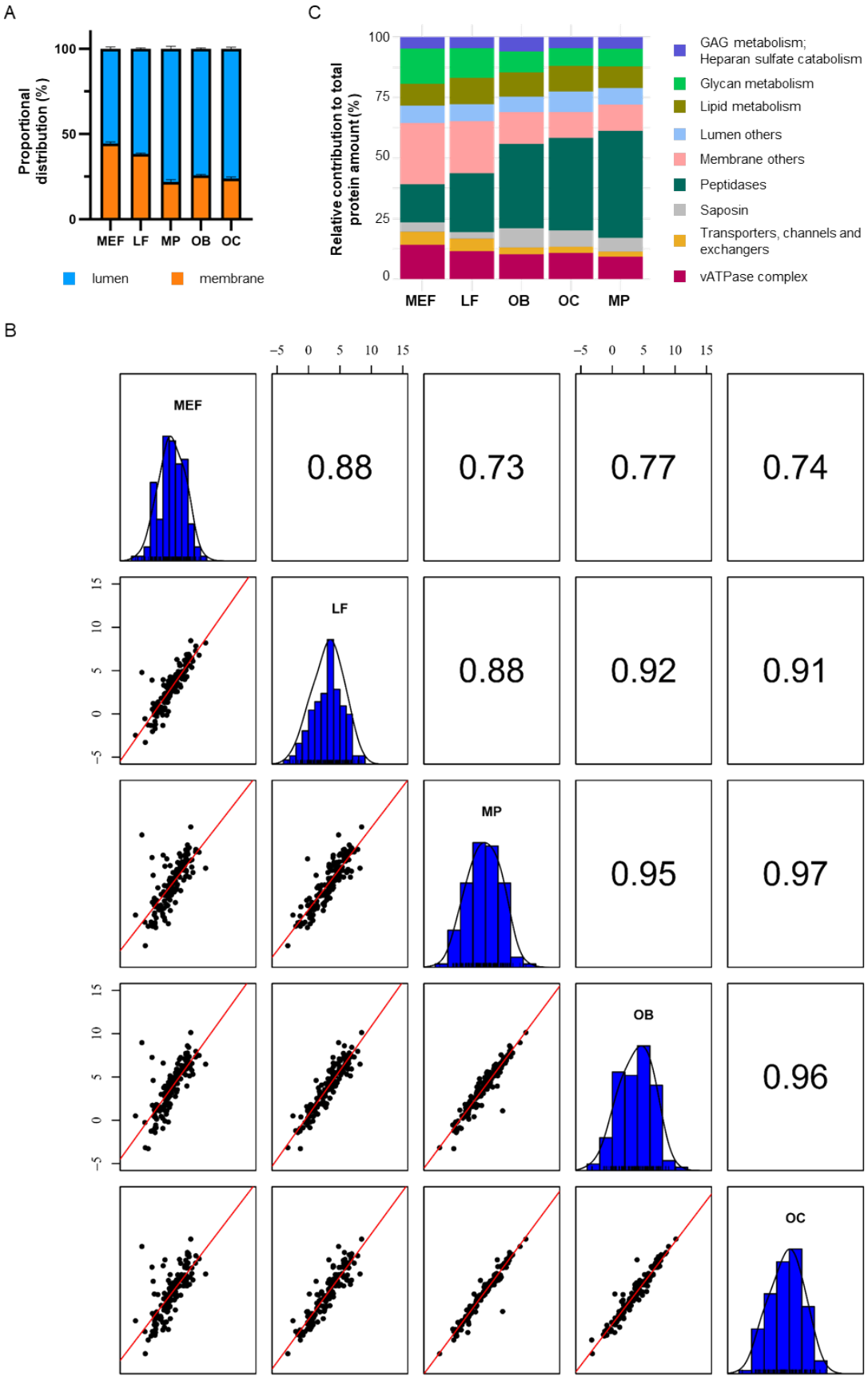


Figure S7: Localization study for lysosomal proteins. A-F) Proportional distribution of individual lysosomal protein molecules (normalized to one lysosome) located in/at the lysosome (lysosomal, left) and outside the lysosome (non-lysosomal, right). A) Each horizontal bar represents a protein. Colors indicate the topology (luminal - blue, membrane - red). Grey horizontal line indicates the threshold cut-off for dual localization. In this analysis, several proteins exhibited values exceeding 100%. Since it is not possible for a protein to be more abundant in lysosomes than in the entire cell, this is likely due to a stronger suppression of individual peptide signals in whole-cell lysates, leading to a underestimation of their abundance. Consequently, these proteins were classified as fully lysosome-localized (see Material and methods, Table S2) Median values across biological replicates (n=3) are reported; error bars indicate the robust standard deviation (rSD).

Figure S8


A
Proportional distribution (%)
100
50
0
MEF LF MP OB OC
lumen
membrane
C
Relative contribution to total protein amount (%)
100
75
50
25
0
MEF LF OB OC MP
GAG metabolism; Heparan sulfate catabolism
Glycan metabolism
Lipid metabolism
Lumen others
Membrane others
Peptidases
Saposin
Transporters, channels and exchangers
vATPase complex
B
MEF
LF
MP
OB
OC
0.88
0.73
0.77
0.74
0.88
0.92
0.91
0.95
0.97
0.96

Continuation Figure S8

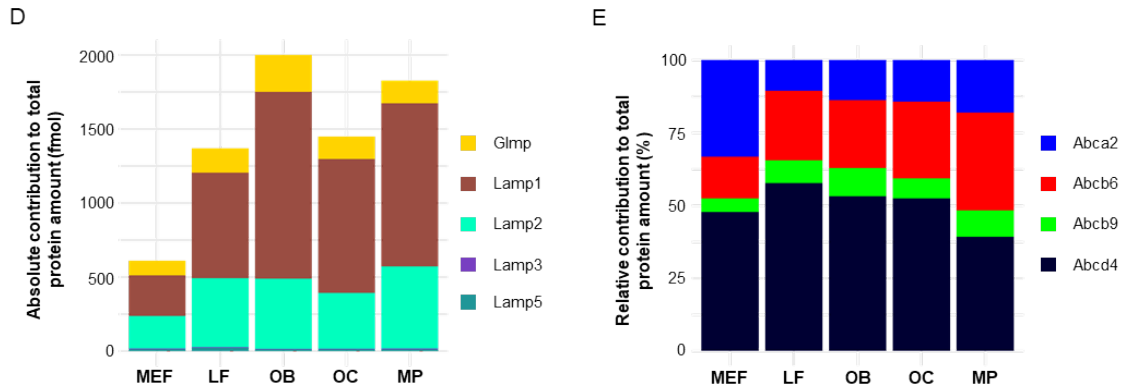


Figure S8: Lysosomal protein expression analysis for individual cell lines. A, C-E) Shown is the relative (%) or absolute (fmol) contribution of individual protein amounts/classes to the total protein amount for A) distinguishing luminal vs. membrane protein localization, C) nine lysosomal functional and structural subclasses, D) structural membrane proteins, E) ATP-binding cassette (ABC) transporters. B) Correlation of lysosomal protein expression profiles. Pearson correlation of log2-transformed absolute protein amounts was calculated. Abbrev.: ABC: ATP-binding cassette, OB: osteoblast, OC: osteoclast, LF: lung fibroblast, OB: osteoblast, MP: macrophage.

8.2 Supplementary information

Supplementary information related to Chapter 3 is provided and comprises:

1) Supplementary file: designed QconCAT DNA constructs

2) Table S1: information on the QconCAT protein standard, generation and characterization

3) Table S2: protein quantification data for MEF and LEF samples as well as dual localization study

4) Table S3: protein quantification for different cell types and correlation between RNA and protein expression