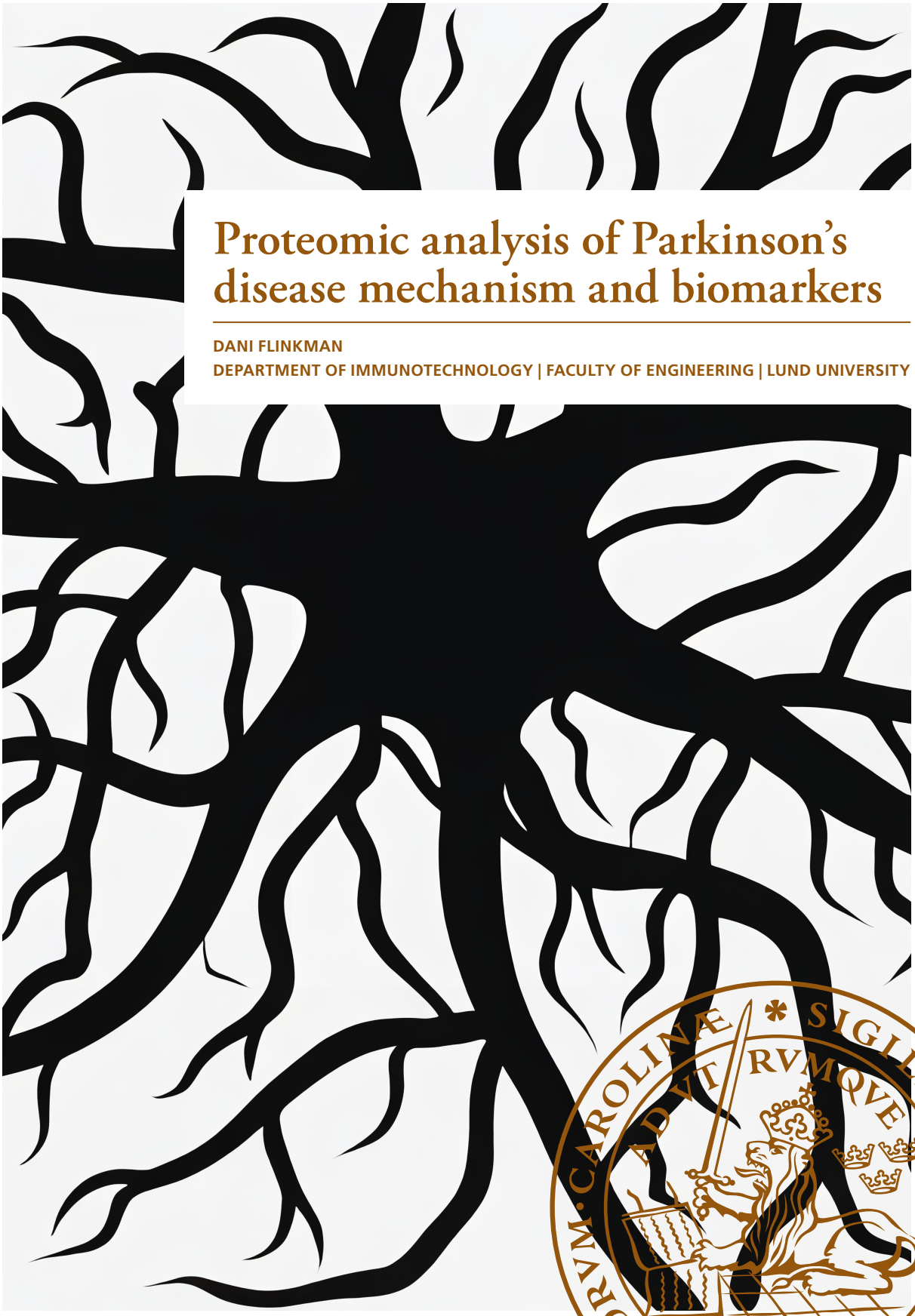
Proteomic analysis of Parkinson's disease mechanism and biomarkers

Flinkman, Dani

2024

Proteomic analysis of Parkinson's disease mechanism and biomarkers

DANI FLINKMAN
DEPARTMENT OF IMMUNOTECHNOLOGY | FACULTY OF ENGINEERING | LUND UNIVERSITY



Proteomic analysis of Parkinson's disease mechanism and biomarkers

Dani Flinkman

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Department of Immunotechnology at Lund University to be publicly defended on 17th of May at 10.00 in Hörsalen, Medicon Village, Scheelevägen 2, Lund

Faculty opponent
Jean-Charles Sanchez

Organization: LUND UNIVERSITY
Department of Immunotechnology
Medicon Village (H406)
SE-22381 Lund, Sweden

Document name: Doctoral dissertation **Date of issue** 2024-05-17

Author(s): Dani Flinkman **Sponsoring organization:**

Title and subtitle: Proteomic analysis of Parkinson's disease mechanism and biomarkers

Abstract:

Aberrant protein kinase signalling is central to multiple neurodegenerative diseases. Increasingly these diseases are investigated with omics techniques with aim to understand changes at systems level opposed to singular signalling or protein level changes events. Proteomics is a field of omics, which aim's to directly measure changes happening in protein level leading to a certain phenotype.
Focus of this thesis lies in Parkinson's disease, which is the second most common neurodegenerative disease affecting millions of individuals worldwide. The most common disease associated mutation in Parkinson's disease is in LRRK2, which is a protein kinase. LRRK2 causes Parkinson's disease phenotype similar to sporadic form of disease making it ideal to study pathogenic mechanisms in this disease.
Work in this thesis revolves around understanding protein kinase signalling in proteomics data. Furthermore, LRRK2 disease causing mechanisms are investigated in animal and patient samples. This revealed that mechanism involved in protein synthesis and proteostasis are dysregulated in both LRRK2 and sporadic form of Parkinson's disease.

Key words:

Parkinson's disease, Kinases, LRRK2, Protein synthesis, Proteomics

Language: English **ISSN and key title**:

ISBN: 978-91-8104-046-3 (print)

ISBN: 978-91-8104-047-0 (electronic)

Recipient's notes **Number of pages**: 94

Price Security classification

Signature Date 2024-04-02

Proteomic analysis of Parkinson's disease mechanism and biomarkers

Dani Flinkman

LUND
UNIVERSITY

Coverphoto by Dani Flinkman

Faculty of Engineering

Department of Immunotechnology

ISBN: 978-91-8104-046-3 (print)

ISBN: 978-91-8104-047-0 (electronic)

Printed in Sweden by Media-Tryck, Lund University

Lund 2024

Table of Contents

Populärvetenskaplig sammanfattning

Avvikande proteinkinassignalering är central för flera neurodegenerativa sjukdomar. Dessa sjukdomar undersöks allt oftare med omics-tekniker i syfte att förstå förändringar på systemnivå i motsats till singular signalering eller förändringar av proteinnivåer. Proteomics är ett område av omics, som syftar till att direkt mäta förändringar som sker i proteinnivån som leder till en viss fenotyp.

Fokus för denna avhandling ligger på Parkinsons sjukdom, som är den näst vanligaste neurodegenerativa sjukdomen som drabbar miljontals individer över hela världen. Den vanligaste sjukdomsassocierade mutationen i Parkinsons sjukdom är i LRRK2, som är ett proteinkinas. LRRK2 orsakar fenotyp av Parkinsons sjukdom som liknar idiopatisk sjukdomsform, vilket gör det idealiskt att studera patogena mekanismer i denna sjukdom.

Arbetet i denna avhandling kretsar kring att förstå proteinkinassignalering i proteomikdata. Dessutom undersöks mekanismer som orsakar LRRK2-sjukdom i djur- och patientprover. Detta avslöjade att mekanismer som är involverade i proteinsyntes och proteostas är dysreglerade i både LRRK2 och sporadisk form av Parkinsons sjukdom.

List of Papers

Paper I

Prasannakumar Deshpande, Dani Flinkman, Ye Hong, Elena Goltseva, Valentiina Siino, Lihua Sun, Sirkku Peltonen, Laura Elo, Valtteri Kaasinen, Peter James, Eleanor Coffey. Protein synthesis is suppressed in sporadic and familial Parkinson's disease by LRRK2. *Faseb journal*, 2020. 34(11): p. 14217-14233

Paper II

Dani Flinkman, Ye Gong, Jelena Gnjatovic, Prasannakumar Deshpande, Zuzsanna Ortutay, Sirkku Peltonen, Valtteri Kaasinen, Peter James, Eleanor Coffey. Regulators of proteostasis are translationally repressed in fibroblasts from patients with sporadic and LRRK2-G2019S Parkinson's disease. *NPJ Parkinson's disease*, 2023. 9(1): p. 20

Paper III

Ye Hong, Dani Flinkman, Tomi Suomi, Sami Pietilä, Peter James, Eleanor coffey, Laura Elo. PhosPiR: an automated phosphoproteomic pipeline in R. *Briefings in bioinformatics*, 2022. 23(1)

Author's contribution to the papers

Paper I

Contributed to conception, organization, design, statistical analysis, execution and writing the first draft.

Paper II

Carried out AHA-labelling optimization, sample preparation and LC-MS/MS DDA, PRM analysis, data pre-processing and contributed to writing the manuscript.

Paper III

Contributed to conception, data preparation and testing.

Abbreviations

αSyn	α-synuclein
AKID	Automatic Kinase-specific Interaction Detection
Ala	Alanine
ANK	Ankyrin
ANOVA	Analysis of variance
APE	Ala-Pro-Glu
Arg	Arginine
ARM	Armadillo
Asp	Aspartate
ATF6	Activation transcription factor 6
ATP	Adenosine triphosphate
AHA	L-azidohomoalanine
ANL	L-azidonorleucine
BONCAT	Bio-orthogonal Non-canonical Amino Acid tagging
CID	Collision-induced dissociation
COR	c-terminal of the Roc domain
DDA	Data dependent acquisition
DFGψ	Asp-Phe-Gly followed by any hydrophobic residue
DIA	Data independent acquisition
DiART	Deuterium isobaric amine reactive tag
DiLeu	N,N-dimethyl leucine
DLK	Dual leucine-zipper kinase
DNA	Deoxyribonucleic acid
EEF2	Eukaryotic elongation factor 2
EGFR	Epidermal growth factor receptor
ePK	Eukaryotic protein kinase
eIF2α	Eukaryotic initiation factor 2 alpha
ER	Endoplasmic reticulum

ERK	Extracellular signal-regulated kinase
FUNCAT	Fluorescent non-canonical Amino acid Tagging
G2019S	Glycine 2019 to serine
GBA	Glucocerebrosidase
GDP	Guanosine diphosphate
Glu	Glutamate
GO	Gene Ontology
GRP78	78-kDa glucose regulated protein
GSEA	Gene Set Enrichment Analysis
GTPase	Guanosine triphosphate hydrolase
His	Histidine
HPG	L-homopropargylglycine
HPRD	Human Protein Reference Database
IBD	Inflammatory bowel disease
IBT	10-plex isobaric tags
Ile	Isoleucine
iMSTIQ	Index-ion triggered MS2 ion quantification
IS-PRM	Internal Standard Triggered-Parallel reaction monitoring
IRE1	Inositol requiring 1
I2020T	Isoleucine 2020 to threonine
HRD	His-Arg-Asp
iBAQ	Intensity-based absolute quantification
iKiP-DB	In vitro kinase-to-Phosphosite database
iTRAQ	Isobaric tags for relative and absolute quantification
JNK	c-Jun terminal kinase
KEA	Kinase enrichment analysis
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KSEA	Kinase-substrate enrichment analysis
LC-MS	Liquid chromatography Mass spectrometry

LFQ	Label free quantification
LOQ	Limit of quantification
LPS	Lipopolysaccharide
LRRK2	Leucine-rich repeat kinase 2
MAPK	Mitogen activated protein kinase
MAR	Missing at random
MCAR	Missing completely at random
MBR	Match between runs
MNAR	Missing not at random
mRNA	Messenger ribonucleic acid
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
MS	Mass spectrometry
MS1	Intact ion scan
LB	Lewy body
Lys	Lysine
N1437H	Asparagine 1437 to histidine
PBMC	Peripheral blood mononuclear cells
PERK	RNA-like endoplasmic reticulum kinase
PINK1	PTEN induced kinase 1
PKA	Protein kinase A
PRM	Parallel reaction monitoring
Pro	Proline
PTM	Post-translational modification
PTMsigDB	PTM signatures database
PTR	Protein to mRNA ratio
R1441C/G/H	Arginine 1441 to cysteine/glycine/histidine
Rab	Ras-associated binding
Ras	Rat sarcoma virus
RNA	Ribonucleic acid

ROC	Ras-of-complex
RP	Reverse phase
RUV-III	Removal of unwanted variation
Ser	Serine
SNpc	Substantia nigra pars compacta
SDC	Sodium deoxycholate
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel
SIL	Stable heavy isotope synthetic peptide
SRM	Selected reaction monitoring
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
Thr	Threonine
TMT	Tandem mass tag
tRNA	Transfer RNA
Tyr	Tyrosine
UPR	Unfolded protein response
WB	Western blot
VAIK	Val-Ala-Ile-Lys
Val	Valine
Y1669C	Tyrosine 1669 to cysteine
Y2018F	Tyrosine 2018 to phenylalanine

1. Introduction

The human brain is a complex organ consisting of multiple functional regions, including dividing cells such as glial and stem cells, cells of blood vessels and about 100 billion neurons. The brain controls all aspects that makes us human: cognition, language, and emotion, as well as deciphering sensory information from outside and regulating movement. This makes brain one of the most complex and important organs in the human body. Neurological disorders are a heterogeneous group of conditions, where the normal function of the brain is disturbed. These disorders range from migraine which affects ~1.04 billion individuals worldwide [1] to more rare disorders such as multiple sclerosis affecting ~2.2 million individuals worldwide [2].

These disorders can severely affect a person's quality of life and at worst leads to a lifelong struggle at a great cost at the level of individual and the society. Taken together the economic cost of these various disorders was estimated to be 798 billion euros in the year 2010 in Europe alone [3]. For many of the disorders there is no cure available, the cause is unclear and the diagnosis leading to treatment can take years. Furthermore, the study of various disorders affecting the brain are hard to carry out. The brain is generally inaccessible to invasive biopsies, animal and cellular models do not always reproduce the human physiology, and post-mortem brains are hard to acquire. Other approaches for studying brain disorders include brain imagining and indirect readouts of brain status from various bio-fluids such as plasma, urine and cerebrospinal fluid (CSF). These are helpful, but do not reveal the molecular mechanisms of the pathology in cells. In this thesis, the focus will be on Parkinson's disease (PD), with special focus on protein kinases and proteomics to study human disease. The disease was first described by James Parkinson in his 1817 description in" An Essay on the Shaking Palsy" and is named after him. Despite this description over two centuries ago and groundbreaking advances in genetics, imagining and biochemistry we have yet to understand the underlying pathogenesis behind the disease.

Advancements in biochemistry and cellular biology have helped to solve how many fundamental physiological and pathological processes work. In the past, much of the research in understanding how these processes work was based on pre-selecting a set of genes and biomolecules of interest and attempting to assign a biological function for them based on a range of experimental approaches. This has led to advancement in understanding basic processes such as glucose metabolism, and

how this process is disturbed in type 1 diabetes, and how insulin can be used to treat this condition. However, many questions remain, which necessitate inspection of large networks made up by the biomolecules found in the cells and other physiological compartments. Understanding the interactions in complex systems - out of necessity- has moved biological research into the omics era. Omics as a term includes the systematic study of large networks of molecules of interest that are present in the cell and living beings. These include studies into small metabolites, polynucleotides, polypeptides, fatty acids and the various modifications made to these. In this thesis, I utilize a subset of omics: proteomics to study human biology and disease.

In **paper I** we studied PD linked Leucine-rich repeat kinase 2 (LRRK2) function in protein synthesis. We show that LRRK2 associates with ribosomal small subunit 40S and that LRRK2 kinase activity supresses protein synthesis in cultured primary hippocampal and dopaminergic neurons, by labelling cells with the methionine analogue azidohomoalanine (AHA) and Fluorescent non-canonical Amino acid Tagging (FUNCAT) approach for quantification. Using same method for protein synthesis estimation, we find that this process is reduced in rotenone and 6-hydroxy dopamine models of PD by 40% and is rescued by LRRK2 inhibition. We next used the rotenone model of PD in vivo in rats, and using phosphoproteomic approach, we found multiple protein synthesis arrest points turned on in striatum and mid-brain. Lastly, we observe attenuated protein synthesis in cultured patient fibroblasts from both sporadic and LRRK2 glycine 2019 to serine (G2019S) forms of PD. This effect is reversible by LRRK2 kinase activity inhibition and is not observed in fibroblasts from multiple systems atrophy patients. This suggests that protein synthesis is dysregulated in PD, and the effect is regulated by LRRK2.

Paper II continued the work from **paper I**. We performed validation of reduced protein synthesis in patient fibroblasts and proceeded to identify newly synthesized proteins that are affected in these cells with Bio-Orthogonal Non-canonical Amino Acid tagging (BONCAT) and bottom-up proteomics. We identified a panel of proteins that are changing specifically in sporadic and LRRK2 G2019S PD, with an overlap of 65% between these two PD types. To validate the markers in PD we used parallel reaction monitoring to see if the changes recapitulate at the total protein level. We find that key regulators of proteostasis are downregulated in both sporadic and G2019S PD. In conclusion, this work identified a panel of proteins that may be involved at down-regulated protein synthesis we observed in **paper I**.

Paper III Mining of large phosphoproteomics datasets and extracting kinase-substrate relations ships from such data, as well as functional prediction, is difficult. We developed a fast and user-friendly tool for phosphoproteomics data analysis in R: PhosPiR. PhosPiR accepts data from popular MS search engine MaxQuant, and data that has been transformed to similar format from other tools. PhosPiR performs offers pre-processing of data and performs multiple statistical tests for finding significantly changing phosphosites. PhosPiR predicts kinase activation from the

data and performs network and functional enrichment analysis. PhosPiR simplifies phosphoproteomic data analysis, which up to date has mostly relied on expert knowledge of the separate tools available.

2. Protein kinases

Protein kinases are a group of enzymes that transfer a phosphate group from adenosine triphosphate (ATP) to specific amino acid residues on proteins in a reversible process called phosphorylation. Protein phosphorylation controls the activity of proteins, localization, and protein-protein interactions. These modifications therefore have profound impact on cellular signalling and function, and not surprisingly, dysregulated phosphorylation is central to many human diseases and conditions. Protein kinases are also highly druggable, which makes them good targets in therapeutics and in the study of cellular processes. As the scope of this thesis is in human diseases, the literature review here is mostly restricted to eukaryotes from Animalia kingdom.

It was believed well up to early 2000s that roughly one third of the proteome was phosphorylated [4, 5]. Technological advances contest these numbers, as modern mass spectrometry (MS) based in depth phosphoproteomic study detected phosphorylation in ~75% out of the roughly 11k detected proteins, in a cell line [6]. As many phosphorylation events trigger under specific conditions, and technological limitations in detection applied, Sharma *et al.* speculate that more than 90% of expressed proteome could be phosphorylated [6]. The most common phosphorylated amino acid residues are serine (Ser), threonine (Thr) and tyrosine (Tyr), which in the 1980s were estimated at a ratio of ~90%, ~10% and ~0.05% of all phosphorylated residues in mammalian cells by autoradiography [7]. A similar ratio was detected empirically using proteomics where the number of phosphorylation events detected was validated [5]. Although not in the scope of this thesis, other phosphorylated amino acid residues are also observed in mammalians such as histidine (His) [8, 9], lysine (Lys) [8], arginine (Arg) [10], and aspartate (Asp) [11]. These are relatively understudied due to low abundance, and due to the fact that basic and acidic residue phosphorylation products are labile under acidic conditions [12, 13], which is standard pH in peptide separation for mass spectrometry analysis. As these are difficult to study, it becomes important to carefully assess the data quality carefully. For example, one study reported over 100 phosphorylated histidine residues under near neutral peptide separation conditions from mammalian cells [14]. Independent assessment of the same data revealed that many of the sites were likely to have been erroneously assigned to histidine [15]. These authors additionally performed an independent experiment, which lead to conclusion that although histidine phosphorylation is genuine in mammalian cells

this is extremely rare, and misinterpreting the data can lead to erroneous over identification of phosphorylated histidine [15]. As very little is known about protein kinases targeting residues other than serine, threonine and tyrosine, eukaryotic protein kinase here refers to protein kinases targeting these three residues.

2.1 Eukaryotic protein kinase structure and diversity

2.1.1 Structure

Understanding the function of protein kinase starts from the protein kinase conserved structure. The first protein kinase structure to be defined was protein kinase A (PKA) [16], which contains typical protein kinase core structure domains found in protein kinases considered typical eukaryotic protein kinases (ePK). These kinases have two conserved major domains in the catalytic core: N- and C-terminal lobes, which together form the catalytic cleft. The N-terminal lobe typically consists of five β-strands and an α-helix called the C-helix. The β-strands are responsible for binding and coordinating ATP for phosphate transfer and contains a conserved glycine-rich loop and valine(Val)-alanine(Ala)-Isoleucine(Ile)-Lys motif (VAIK) necessary for proper ATP positioning [17]. The C-helix is positioned for efficient catalysis in an active kinase [17].

The C-terminal lobe is the catalytic core of ePK consisting of α-helixes and a β-sheet consisting of four β-strands, and is responsible for substrate binding and phosphate transfer [17]. This domain contains the kinase's activation segment, which functions as an on/off switch for kinase activity and binds to the phosphorylation region of the substrate. This segment is highly variable amongst kinases and starts with a conserved Asp-phenylalanine(Phe-)-glycine(Gly) followed by any hydrophobic residue motif (DFGψ) and extends 20-35 amino acids up to Ala-proline(Pro)-glutamate(Glu) motif (APE) [18] (Figure 1.). The DFGψ motif is responsible for magnesium binding necessary for proper ATP positioning, while the APE motif helps to dock activation segment motifs to F-helix, which in turn helps to stabilize the active conformation of the kinase[17].

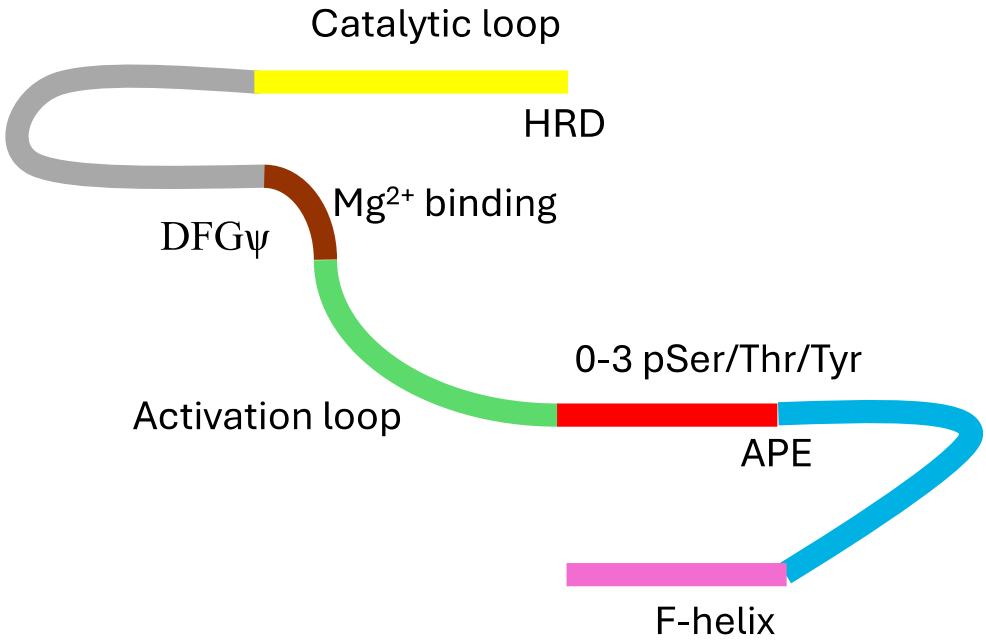


Figure 1 Schematic view of the protein kinase activation loop.

The activation segment contains the activation loop, which in some kinases is auto phosphorylated or phosphorylated by another kinase at one or more Ser/Thr/Tyr residues. Phosphorylation of these residues, which are 5-10 amino acids N-terminal to the APE sequence drives the activation of some kinases by docking the phosphorylated residue to the arginine of His-Arg-Asp (HRD) motif located in catalytic loop, as well as to C-helix and residues inside the activation loop [17]. Kinases that do not use this mechanism have an amino acid sequence in activation segment that form the active conformation with HRD motif without phosphorylation [18].

Adjacent to the activation loop is the P+1 loop, which binds to the substrate sequence C-terminal to phosphorylation site. In the activated kinase this region binds a region of the substrate which tends to have disorganized conformation surrounding the phospho-acceptor site [19]. Although this segment contributes to substrate specificity, the substrate and scaffolding protein binding grooves outside the catalytic cleft also contribute to specificity. For example c-Jun terminal kinase (JNK) family member 1 has a C-terminal docking groove for substrate and docking

proteins [20], and PD associated PTEN induced kinase 1 (PINK1) has three unusual N-terminal inserts, where the third insert helps to create a binding motif for ubiquitin [21]. Active kinases undergo an open-close cycle, where correctly folded and active kinase catalytic cleft closes on the substrate and orients the target residue catalytic loop HRD motif aspartate and ATP. Following the catalysis, the catalytic cleft opens, and the phosphorylated substrate is ejected from the cleft, which allows the kinase to phosphorylate next substrate.

As said earlier, the catalytic cleft is between the C- and N-terminal lobes, and therefore it is not surprising that the amino acid residues from these domains interact and form conformational motifs. In an active kinase two leucine residues from the N-terminal lobe interact with Phe from DFGψ motif and His/Tyr from HRD motif forming a hydrophobic regulatory spine, which is called the R-spine in the literature, the formation of which is the ultimate indication of an activated kinase [22]. During activation the second spine forms and is referred as C-spine standing for catalytic spine. In PKA C-spine is formed by Val from β2 strand and Ala from VAIK motif, and Leucine 173 from C-terminal lobe dock with adenine ring from ATP forming a catalytic spine referred to as C-spine in literature, which servers to position ATP correctly [23]. Both of these spines are anchored to the F-helix C-terminal to activation segment ensuring an ordered structure necessary for catalysis [17]. The structures of over 300 kinases show a high level of conservation in an active state, while the inactive state shows more variation [24]. Consistent with these spines being highly conserved in and necessary for active kinase conformation, amino acid residues in the spines are mutually exclusive with residues that determine kinase substrate specificity [25].

2.1.2 Kinome

The exact number of protein kinases that make up the human kinome is unknown. Peer-reviewed evidence stands at 525-550 serine, threonine, tyrosine protein kinases [26, 27] with nine major subfamilies based on sequence similarity [26]. Protein kinase classification for many kinases relies on sequence and structural homology predictions, as well as substrate and inhibitor specificity. However, many proteins have been overlooked for their phosphate transfer activity towards proteins, and it is likely that there are more than in the reviewed literature [28, 29]. For example, some non-protein kinases show dual specificity for proteins and small molecules [30] and some pseudokinases are catalytically active and will be discussed later. According to the most recent classification by Moret *et al*., proteins exhibiting protein kinase activity can be divided in to 3 classes, and that way expand the definition of kinome to 710 protein kinases [29]. These classes are ePK, eukaryotic like kinases that differentiate from ePK in substrate binding lobe, and finally atypical kinases that adopt kinase like 3D-structure, but have weak sequence similarity to classical protein kinases [29]. Kinase terminology can cause confusion,

as same terms tend to have different definitions between studies hindering the research. For example, vascular endothelial growth factor receptor is a tyrosine kinase that has four alternative names and four different gene names in the literature according to UniProt [31], and only three names include the term kinase. With the introduction of the AlphaFold database [32] with predicted structures for all proteins, it is likely that more proteins with kinase like 3D structures will be identified, furthering the endeavour to find more proteins exhibiting protein kinase activity.

Some kinases classify as pseudokinases, which number at roughly 50. Pseudokinases have high sequence homology to catalytically active kinases, but lack key amino acids thought to be indispensable for catalytic activity. These amino acids are lysine in the VAIK motif involved in ATP binding, the aspartate catalytic loop's HRD motif and the DFGψ motif involved in Mg binding [33]. These kinases lack catalytic activity, and yet they are thought to be active components of signalling cascades by interaction with other kinases, by sequestering or anchoring kinase substrates or serving as signalling scaffolds [33]. While these motifs and residues are indispensable for kinase activity there is evidence that at least, some these so called pseudokinases are active. For example, CASK [34] and Haspin [35, 36] are both catalytically active. Both lack canonical DFGψ motif, and Haspin does not contain an APE motif. It is possible that the definition of a pseudokinase is due to lack of thorough characterization. For example, PD associated Leucine-rich repeat kinase 2 (LRRK2) has an unusual DFGψ motif, which serves to keep the kinase in inactive form [37], but was originally discovered as an active kinase. Therefore, it remains a distinct possibility that an additional number of, non-protein kinases, uncharacterized proteins and pseudokinases can also phosphorylate proteins, but their mechanism of action and binding partners and substrates are simply not known due to lack of studies or appropriate screens.

2.1.3 Kinase signalling

Kinase signalling is ubiquitous in most cellular processes and dysregulated signalling can lead to many diseases including cancer and neurodegeneration. Kinase signal transduction often happens in multiple phases, and the stimulus can lead to multiplication of the original signal. A signalling cascade usually starts by a certain stimulus, such as binding of a ligand to a receptor, accumulation of small molecules such as Ca^{2+}, and unfolded proteins, or nutrient deprivation. A classic example of protein kinase signalling through a cascade are mitogen activated protein kinases (MAPK), key regulators of cell survival, growth and death. Extracellular signal-regulated kinase (ERK) signalling regulates various responses coming from the outside of the cell and is an example of typical MAPK signalling. Here an extracellular ligand binds to an tyrosine kinase receptor, which leads to activation of a Rat sarcoma virus (Ras) guanosine triphosphate hydrolase (GTPase), which

triggers activation of upstream MAP3K kinase, which activates MAPK2 by phosphorylation, which in turn proceeds to activate ERK by phosphorylation [38]. Organisation of kinases within a cascade imposes a threshold for activation that ensures that the cascade of events do not occur unless a sufficiently strong trigger is present. The organisation of MAPK cascades with each consecutive level expressing more molecules, facilitates a high level of signal amplification [39].

Many of the kinase signalling pathways have overlapping activators, substrates, and intermediates. Additionally, multiple signalling pathways can cross talk and integrate into a specific response. As an example dual leucine-zipper kinase (DLK) is an upstream activator of both the JNK family kinases and p38 [40]. As an upstream MAPK3-level kinase, DLK regulates multiple processes from cell degeneration and death to regeneration, and is able to activate both JNK and p38 MAPK pathways, which can cross-talk in specific context such as during axon regeneration [40]. As aberrant kinase signalling can cause serious consequences, kinase regulation is achieved in multiple ways such as post-translational modifications, subcellular localization and by ensuring that correct proteins interact at the correct time.

Regulation by localization is achieved in multiple ways. Some protein kinases have a localisation tag, for example, unprocessed PINK1 has a mitochondrial localization tag. Many kinases are trans-membrane proteins localised at the plasma membrane and some kinases are specifically localized to subcellular localizations due to protein interactions. Maintaining the correct localization is an important function for physiology and cell death, and a wrong localization can trigger incorrect signalling outcomes. As an example, members of the JNK kinase family come from three genes and regulate a wide variety of functions from liver toxicity to neuronal development [41]. JNK family kinases have some overlapping substrates such as stathmin-2, which is preferentially phosphorylated by JNK1, while JNK2 and JNK3 phosphorylate it with less efficiency [42]. JNK1 is physiologically active member of JNK family, while JNK2 and JNK3 are stress activated [43, 44]. As these kinases have substrate overlap, it is important to regulate where and with what JNK1 interacts with, to avoid erroneous stress response without a trigger, which is more the domain for JNK2 and JNK3. In dorsal ganglion root sensory neurons, JNK2 and JNK3 are palmitoylated, but not JNK1 [45]. This modification allows stress induced localization and activation of JNK3 by DLK, but not other MAP kinases including JNK1, which lack the palmitoylation modification [45]. Palmitoylation of JNK and DLK [45] ensures that in the case of cell stress, correct proteins are in the correct place, able to interact and generate the correct outcome.

Kinase signalling can create a positive feedback loop, or the kinase can inactivate itself. For example activated proteins in the ERK pathway ERK can directly inactivate their activators by phosphorylation [38]. As an example, activation of ERK leads to phosphorylation of the epidermal growth factor receptor (EGFR) at Thr669 by ERK. This disturbs EGFR dimerization and cross-activation, leading to

attenuation of EGFR-ERK signalling [46]. If the negative feedback loop for ERK signalling is not working, the consequence is aberrant cell signalling and can lead to oncogenesis.

As the specific outcome from a kinase depends on many factors, it is important to take a systems approach to understand molecular events happening in cell, and to understand key signalling drivers. In the next chapter I will discuss methods to study kinases biochemically and in systems biology.

2.2 Studying eukaryotic protein kinases.

2.2.1 Biochemical methods for studying kinase function

Various biochemical methods exist to study protein kinases and some commonly used methods are discussed briefly below. One of the most widely used methods to measure protein phosphorylation is Western blotting (WB) [47]. In WB proteins are separated on a sodium dodecyl sulphate polyacrylamide gel (SDS-PAGE) [48] and transferred to a membrane, and subsequently proteins are detected with an antibody recognizing the target protein. Antibodies detecting a specific phosphorylated residue are in routine use in many laboratories and provide information on single phosphorylation events. This approach is valuable in detecting well-characterized signalling events such as detecting kinase activation by autophosphorylation or by substrate phosphorylation. However, this method is low throughput and impractical for understanding systems level changes and the integration of signalling pathways. Antibodies are also prone to unspecific binding [49], and according to a 2015 estimate, almost half of the money spent on antibodies is spent on antibodies that do not properly recognize their targets [50]. Phosphorylation-specific antibodies have another disadvantage; neighbouring phosphosites can change the affinity of the phosphosite specific antibody [51], which makes thorough antibody characterization or complementary evidence important in kinase studies using antibodies.

SDS-PAGE gels also have other uses, such as to detect phosphorylation events in mechanistic studies where antibodies are not used. Classical methods use [γ-^{32P}]-labelled ATP and perform a kinase reaction *in vitro* with purified kinase and substrate, and separate the reaction components on the gel and detect phosphorylation by autoradiography [52] This is useful in finding and validating substrates, testing efficiency of molecules affecting kinase activity as, well as studying mutations that might affect kinase activity. Kinase activity can also be detected in non-radioactive gel assays by looking at the protein mobility shift on a gel. If the charge density between phosphorylated and non-phosphorylated forms of protein is large enough, the protein migration on gel will be altered [53].

Alternatively it is possible to induce a phosphorylation-dependent retarded migration on SDS-PAGE gel by adding a reagent known as “Phos-tag” to the gel to distinguish the phosphorylated protein from non-phosphorylated[54].

Functional validation of phosphorylated proteins is possible with point mutations and kinase inhibitors. Using mutations involves that either the kinase or the substrate has the phosphosite of interest mutated in to phosphomimetic amino acid glutamate and Asp to simulate phosphorylation, or using inhibitory amino acids such as Ala or glutamine for Ser/Thr residues or Phe for Tyr residues to prevent phosphorylation [55, 56]. These mutations have multiple purposes. They can serve as internal controls for substrate screening and, they can be used to provide functional information in biological systems *in vivo* and *in vitro* on consequence of phosphorylation. For example, JNK1 knockout mice have abnormal neurodevelopmental phenotype, which is rescued by mutating JNK specific Stathmin-2 phosphorylation sites to aspartate, which mimics the JNK phosphorylated form of Stathmin-2 [57].

As mentioned earlier protein, kinases are highly druggable, and several types of inhibitors exist. Type I inhibitors compete with ATP, and stabilize the kinase in an active closed conformation, where the DFGψ motif and C-helix are oriented inwards to catalytic cleft and type I½ inhibitors bind inactive conformation with DFGψ motif facing inwards [58]. Type II inhibitors bind kinases in inactive conformation with DFGψ motif facing out [58]. Type III and IV inhibitors are both allosteric modulators, where type III binds next to the ATP binding pocket, and IV binds outside the substrate and ATP binding pocket [58]. Type V inhibitors bind two distinct sites on the kinase, while type VI covalently bind and inhibit the enzyme [58] and often target free cysteine residues in catalytic cleft [59]. Furthermore, the inhibitors can be divided into pharmacological inhibitors, and protein and peptide derived inhibitors. Many inhibitors are not strictly specific to their targets, and ideally, studies should utilize multiple inhibitors. Additionally using different types of inhibitors can yield mechanistic information, as kinases can change localization and binding partners depending on their conformation, which are changed by the inhibitors.

Ultimately as protein kinases function in large dynamic networks, that have complex cross talk, it becomes increasingly important to understand kinase-signalling networks as a whole. The most systematic and informative way to do this is by proteomics and systems biology methods. Currently mass spectrometry-based proteomics is the most mature technology for systematic study of phosphorylation. Typically, this is done by cleaving proteins into small peptides, which are readily analysed by Mass spectrometers. The exact details of which will be discussed later in chapter 4. However, special biochemical features of phosphorylated dataset generation are discussed here. Peptide sequences containing phosphorylated residues make up a minority of the generated peptides and are hard to detect from the mixture, which contains non-phosphorylated peptide. Therefore, it is

advantageous to enrich phosphorylated peptides instead of proteins. Peptides containing these post-translational modifications (PTM) can be enriched with strong cation exchange [60], strong anion exchange [61], immobilized metal ion affinity [62] or titanium dioxide enrichment [63]. Combination of orthogonal techniques such as strong cation exchange with immobilized metal ion affinity or titanium dioxide enrichment can be used to improve depth of the analysis or for further enrichment [5, 64]. Although all phosphorylated peptides are enriched with these techniques, only a small proportion contain phosphorylated tyrosine. To circumvent this problem, antibodies recognizing pTyr residues or specific pTyr motif binding domains are used to enrich phosphorylated tyrosine containing and peptides [65]. In the next section, I will discuss some of the computational methods used to study protein kinase signalling data derived from MS-based proteomics experiments.

2.2.2 Computational methods

Protein kinase phosphorylation site databases

Many of the systems level phosphorylation methods rely on phosphorylation databases. These databases allow one to make predictions on the kinase activation, data, and functional significance of the phosphorylation sites in the data. These have been recently reviewed extensively by Zhao et al., and although outside the scope of this thesis to review all of them [66], it is good to note specific features of some of the databases. The disadvantage of all databases is that many kinases remain understudied, and ca. 90% of known kinase-substrate relations belong to 20% of protein kinases [67]. The problem is not limited to study of protein kinases, but to the field of biological research as a whole. Some genes are very extensively studied, which amounts to the 100 top most studied genes being present in 25% of biomedical research articles [68]. Furthermore, roughly, 20% of proteins in humans do not have any known function [69] and at least part of the remaining 80% are not fully characterized.

Most extensive databases are those, which contain data from high throughput studies with minimal manual curation, and examples of these are PhosphoSitePlus [70] and Peptide Atlas [71]. PhosphoSitePlus as of writing this thesis contains ~294,000 detected phosphorylation sites Ser/Thr/Tyr in human, mouse and rat, and also data from low throughput experiments from the literature [70]. Similar to PhosphoSitePlus databases combining low and high throughput data, some databases are species-specific, exemplified by PhosphoGRID for S. cerevisiae [72]. This wealth of data generated by high throughput experiments is not without problems as only 5% the known human phosphorylation sites have a known function or assigned kinase [67]. Thus, interpreting high throughput phosphorylation data now heavily relies on specialized downstream tools to predict kinase-substrate relationship. Furthermore, many of the uncharacterized sites

identified in high throughput studies are coming from low quality MS2 peptide to spectrum matches. A consequence of this is that many of these sites are identified in only one study, possibly indicative of false positive identification, which can be problematic [73].

Some databases such as Signor [74, 75], PTM signatures database (PTMsigDB) [76], Human Protein Reference Database (HPRD) [77] and BioGrid [78] specialize in manually curated information on kinase-substrate relationships, PTMs and functional consequence of the PTMs. PTMsigDB, also contains information from PhosphoSitePlus [70], and the information has been added to PTMsiDB manually or semi-automatically. One problem with the manually curated data is the fact that manual data mining and annotation is enormous task even for large collaboration networks. For example, a well-known LRRK2 autophosphorylation site at Ser1292 is not entered in HPRD, BioGrid and Signor (June 2023), while PTMsigDB contains the site owing to its semi-automated information gathering. Tools and databases for assigning kinases to uncharacterized sites are described below.

As mentioned earlier, very few phosphosites have an assigned kinase. Several tools exists that can be used to predict kinase-substrate relationships based on properties such as sequence motifs, 3D-structure and protein-protein interactions, and are reviewed by Savage and Zhang [79]. Many of the tools are no longer being maintained and I will discuss two examples of currently maintained tools below, with distinct functions. The simplest form of prediction uses the sequence motif. However, many motifs have non-unique features, and if considered alone, tens of kinases can map to a single phosphorylation site. To circumvent this problem a tool named NetworKIN combines sequence motif and protein interaction network analysis on the observed phosphorylation sites to predict a kinase for the phosphorylation sites [80]. This tool is useful for building library for well-characterized kinases. According to year 2019 estimate by Needham et al. 276 protein kinases had five or less substrates, and remaining 149 described in the study had no known substrate [67], and this puts limits on a known sequence motif based kinase prediction. An alternative approach proposed by Parca *et al.* also uses sequence information, but instead of known sequence motifs, the approach termed Automatic Kinase-specific Interaction Detection (AKID) uses sequence information from both the protein kinase and substrate without prior data [81]. AKID first identifies the protein kinases from sequence with a Hidden Markov Model, and then determines the kinase's specificity conferring residues with KINspect algorithm [25]. Following this, AKID uses deep neural networks to train and detect kinase-substrate relationships using the KINspect output and known kinase-substrate relationships from PhosphoSitePlus and PhosphoGrid. AKID outperformed other prediction methods including NetworKIN but has not been cited in an independent comparative study.

Many kinases have shared features in their substrate phosphorylation motif sequences [82, 83]. These shared features cluster significantly within protein kinase

families [83], and are possibly sites of PTM cross talk [83, 84], conferring specificity and dynamic regulation for substrate recognition. As the substrate phosphorylation motifs are shared between many kinases, special tools are required to assess kinase activation, and will be discussed below.

Assessing kinase activation from phosphoproteomics data

Kinase activation can be assessed with kinase activation prediction tools as described above. There are many tools available, which are reviewed by Piersma *et al.* [85]. Here I will summarize some tools, with their distinct features. One of the earliest still currently maintained computational implementations of predicting kinase association in the dataset is kinase enrichment analysis (KEA) [86, 87], which as of writing this utilizes both predicted and known kinase-substrate interactions. KEA assesses kinase activation by comparing user input to KEA background libraries by performing Fisher's exact test. This kind of analysis does not consider intensity values, and the input is a simple list of proteins. The resulting output does not tell if a particular kinase is activated or inactivated. This aspect is considered in popular kinase-substrate enrichment analysis (KSEA) [88, 89], which to date is the most cited kinase activity prediction tool [85]. KSEA uses prior knowledge from PhosphoSitePlus and predicted substrate sites from NetworKIN. Kinase activation is scored with a modified z-score. The statistical significance of kinase activation is subsequently obtained with one tailed paired Student's t-test and corrected for multiple testing with Benjamini-Hochberg method. KSEA is simple and easy to understand, which is probably why it continues to be cited.

Another approach presented in PTM-SEA [76] utilizes a popular algorithm for functional enrichment analysis: Gene Set Enrichment Analysis (GSEA) [90]. In simple terms, GSEA first calculates a combined score from p-value and fold change and ranks genes according to combined score so that positive values will be at top and negative at bottom, indicative of up and downregulated genes respectively. Following this GSEA calculates an enrichment score from top to bottom so that whenever a gene is present in a gene set, the score is added to a running sum, and when it's not the score is subtracted, effectively indicating if a gene set is up or down regulated, followed by statistical significance analysis and control for multiple testing. Although GSEA was originally developed for gene set enrichment analysis, this can be utilized with other data, including large phosphorylation datasets. Instead of gene sets, PTM-SEA uses phosphorylation pathways from PTMsigDB [76], which contains the direction of change for each PTM in a pathway. PTM-SEA uses this database to perform modified single sample GSEA to predict kinase activation and inactivation [76]. PTM-SEA can be considered a step up from both KEA and KSEA, but it is still limited by the information present in PTMsigDB, which is derived from known biological and kinase-substrate information, which are limited for most kinases.

To circumvent this bias Mari et al. proposed the In vitro kinase-to-Phosphosite database (iKiP-DB) to predict kinase activation [91]. This approach predicts kinase activation with a dataset coming from a high throughput in vitro kinase substrate screen study with 354 recombinant protein kinases [92]. The raw data from the kinase screen was reanalysed by iKiP-DB authors to yield 160k kinase to phosphosite relations in 4,032 protein groups for 313 kinases. These kinase to phosphosite relations are used to predict kinase activation with PTM-SEA algorithm with expended kinase set relative to PTMsigDB [91]. One disadvantage of iKiP-DB is that the data is coming from an In vitro screen, and many of the substrates might not be real substrates for a given kinase in vivo, which is also noted by the authors [91].

Assessing kinase function in phosphoproteomics data

The simplest way to make predictions of functional changes in data set is to look at functional term enrichment. Popular tools for this include Gene Ontology (GO) terms [93, 94], Kyoto Encyclopedia of Genes and Genomes (KEGG) [95], WikiPathways [96, 97] and Reactome [98]. These tools enable the identification of enriched functions among the significantly changing phosphorylation sites, and some tools like KEGG, Reactome and WikiPathways provide manually curated pathway maps. Although the maps are simplified abstractions, these help to visualize signalling cascades. The downside of these tools is the reliance on known pathway information, which might not originate from protein level experiments and the low amount of functional information for some proteins.

Identifying uncharacterized signalling pathways can take advantage of known and predicted protein-protein interactions, and several tools exists to do this kind of analysis for phosphoproteomics data such as Signor [74, 75], Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) [99, 100], Phosphomatics [101] and human tissue specific tool TissueNet [102, 103] and HumanBase [104]. STRING is one of the most popular protein pathway enrichment and network tools, and integrates known interactions from multiple databases such as HPRD, GO, KEGG, BioGrid and Reactome, with predicted interactions from literature text mining, gene and protein level co-expression [99, 100]. STRING also allows the visualisation of interactions, which is useful for deciphering signalling pathways. Several phosphorylation specific network tools exist. As an example, PhosphoPath builds interaction networks using PhosphoSitePlus, BioGrid and WikiPathways to visualize quantitative changes in top enriching kinase pathways [105]. PhosphoPath also differentiates between phosphosites that have been detected either in peptide without other phosphorylated residues, or in multi-phosphorylated peptides. Phosphomatics is a web based protein phosphorylation network analysis tool, which uses information from Signor [74, 75], UniProt [31], KEGG [95] and BioGrid [78] to annotate the phosphorylation data [101]. Phosphomatics can also draw kinase-substrate relation networks from data; plot the phosphorylation changes on KEGG

pathways. Furthermore, phosphomatics predicts kinase-substrate relationships with known interactions from BioGrid and matches unassigned sites to known kinase-substrate relationships using phosphosite sequence information.

Many of the protein-protein interactions are likely to be tissue and cell type specific, so matching all known interactions from all human cell types to a neuron might mask the enrichment of real interactions occurring in neurons. Tools like TissueNet [102, 103] and HumanBase [102, 103] are specifically made for tissue specific interaction analysis, while some databases such as Signor [74, 75] offer some tissue-based functionality. Although tissue/cell specific interactions are not widely taken into account, it is conceivable that a phosphorylation event in JNK1 in brain might have different consequences to the same event in intestine due to different set of JNK1 interacting proteins being present. Another commonly overlooked issue with functional analysis is the subcellular localization of the phosphorylation events which can be monitored [106, 107], but as this kind of data is not routinely collected and no published tools for subcellular phosphoproteomics analysis are published, it will not be discussed further.

Phosphoproteomic data pipelines

Several of the tools discussed above also function as a part or as a specialized phosphoproteomics pipeline. The aim of these pipelines is to automate data processing and bring high throughput phosphosite analysis tools to bioinformaticians and non-bioinformaticians alike in one package. Each of the pipelines presented here have their own design choices that come with their own limitations and advantages. I will describe some recently released pipelines below.

PaDuA is a phosphoproteomic data analysis pipeline built as a Python package [108]. PaDuA performs sample median subtraction as normalization step, and for imputation the user chooses between random sampling from a normal distribution function similar to that used in Perseus analysis software [109] or an estimate imputation value from similar phosphosites as described by Webb-Robertson, B.J. et al. [110]. PaDuA uses Student's t-test and analysis of variance (ANOVA) for statistical analysis, and PhosphoPath [105] and KEGG [95] for pathway enrichment analysis and visualization [108]. PaDuA accepts output from MaxQuant [111], which is a popular MS/MS data analysis pipeline, or data that has been converted to MaxQuant format. The advantage and limitation of PaDuA is that it is a relatively simple package that does not incorporate a large amount of information. Although this is also a limitation, it makes data interpretation easier.

PhosPiR is a R language-based data analysis pipelines [112] (Paper III). The users can use their own imputation or normalization methods, or use proBatch [113] and/or MSimpute [114] packages offered by PhosPiR for normalization and imputation. For statistical tests PhosPiR offers the choice of Student's t-test, Wilcoxon signed-rank test, reproducibility-optimized test statistic [115] and rank

product test [116] for two-sample statistical analysis, and ANOVA and linear mixed model for unpaired and paired multiple sample test, respectively. PhosPiR uses GO, KEGG, WikiPathway and STRING for protein functional enrichment and network analysis for proteins. For statistically significant phosphosites, PTM-SEA is used for kinase activation and pathway enrichment. Additionally PhosPiR performs a kinase activity prediction with KinSwingR [117], in which PhosPiR uses PhosphoSitePlus kinase-substrate relationship database, and performs a consensus sequence motif analysis on the input data. Subsequently KinSwingR combines the direction of fold change and p-value for a given kinase and calculates p-value for kinase activity change. Finally, PhosPiR visualizes the top 250 phosphorylation sites and their predicted upstream kinases in a Circos plot. A special feature of PhosPiR is a sequence alignment of the phosphosites to a human ortholog, which is then annotated with UniProt information for non-human organisms. PhosPiR accepts phosphosite output from MaxQuant and other output formats need to be converted to this.

PhosR is an R language-based phosphorylation data analysis pipeline and accepts data from MaxQuant and from various commercial MS/MS analysis software packages. PhosR places special emphasis on data imputation and normalization. PhosR performs imputation before data normalization, by performing a similar imputation function as Perseus [109], first in condition-specific way if the phosphosite has a greater value than the user set threshold of x% of non-missing values, and remaining missing values are imputed from total data. Finally, PhosR normalizes the data by removing unwanted variance method [118] with minimally changing phosphosites across samples. Similar to KSEA and KinSwingR, PhosR takes sequence motif and fold changes into account, and scores for top changing kinases in the dataset. PhosR also visualizes protein modules, which consists of proteins with similar kinase regulation profile, which allows for finding proteins that are regulated in a similar way in multiple phosphorylation sites.

3.Parkinson's disease

3.1 PD risk factors and pathophysiology

Parkinson's disease is the second most common neurodegenerative disease, classically characterized by loss of dopaminergic neurons in Substantia nigra pars compacta (SNpc), and its motor symptoms. PD affected 6.1 million people globally in 2016 [119]. This number is growing rapidly and was estimated to have increased to 8.5 million affected by year 2019 [120]. At the population, level the disease incidence increases with age. In a 2016 meta-analysis of incidence studies, PD incidence starts at 3.57/100,000 in males and 3.26/100,000 in females from ages 40-49 [121]. This progressively increases to 258.47/100,000 in males and 103.48/ 100,000 in females at age of +80 [121]. Significant regional variance in incidence rate was reported between the studies reflecting different populations being comparted [121]. The incidence has increased in most world regions between 1990-2019 [120], and it is unclear how much of the increase can be attributed to globally increased life expectancy, aging societies and improved awareness and diagnosis of the disease.

Most prominent and well-known features in PD are the classical progressive motor symptoms, which are many and not limited to: bradykinesia, resting tremors, rigidity, postural instability, and gait impairment. These symptoms can be attributed to progressive degeneration of dopaminergic neuron axons in the striatum followed by loss of dopaminergic neurons in SNpc [122, 123], which is a process that starts before manifestation of motor symptoms and clinical PD diagnosis. SNpc dopaminergic neurons are not the only neurons affected in the brain, and loss of neurons can be observed in multiple brain regions [124]. Increasingly the non-motor symptoms are also being recognized as important features in PD and can precede the motor symptoms and diagnosis of PD by years. These include, but are not limited to: constipation, sleep disturbances, olfactory dysfunction, daytime sleepiness and depression [124]. In respective to these symptoms PD is heterogeneous and there have been attempts to divide PD into subgroups [124], for example with clinical information [125, 126] or biomarkers [127]. A recently proposed staging system classifies PD based on presence of pathological α-synuclein (αSyn), genetic genetic status and clinical features [128].

Currently the cause of the disease is not known. Risk factor for PD include genetic variation, which are discussed later, as well as non-genetic factors such as rural living and exposure to pesticides [124]. Reduced risk for PD is associated with tobacco smoking, alcohol consumption, and coffee drinking [124]. There is no cure for the disease, and treatments aim to slow the progression of the disease and easing the symptoms. There is interest in diagnosing the disease in pre-motor phase, which could lead to significant improvement in life quality for people at risk for PD [129], but currently the disease is most commonly diagnosed at the time when motor symptoms manifest as the prodromal non-motor symptoms are vague.

As discussed above pathologically PD is characterised by loss of dopaminergic neurons in substantia nigra of midbrain and to lesser degree in other neurons in other regions. Many of the surviving and dying neurons also exhibit appearance of protein aggregates called Lewy bodies (LB). One component of these protein aggregate bodies is αSyn, which gene's mutations and gene multiplications are known to cause PD [130-132]. Below I will be discussing LRRK2, and some protein pathways related to PD, with special emphasis on LRRK2.

3.2 LRRK2

LRRK2 is a large ~286 kDa serine/threonine protein kinase and it's mutations are associated with autosomal dominant form of familial PD [133-138], sporadic PD [138] and Crohn's disease [139]. Gain of function mutation in the kinase domain Glycine 2019 to Serine (G2019S) is the most common PD associated mutation and present in 1-6% in sporadic PD [140] and up to 40% in familial PD depending on the population [141]. LRRK2 PD resembles sporadic PD, though the symptoms and disease progression are milder [142]. The symptoms are almost indistinguishable between LRKR2 and sporadic PD. Furthermore, LRRK2 activity is elevated in sporadic PD [140, 143-145], and this makes LRRK2 one of the most promising targets to study in PD biology, and unsurprisingly multiple clinical trials are currently targeting LRRK2. The fact that LRKR2 mutations cause symptoms similar to sporadic PD, we chose LRRK2 as our gene of interest (Paper I and II). LRRK2 is widely expressed most prominently by cells of lung, kidney immune system and the gut, as well as glial cells and neurons [146, 147].

LRRK2 forms a dimer under physiological conditions [148], and contains two catalytic domains: a GTPase domain consisting of Ras-of-complex (ROC) and the C-terminal of the Roc domain (COR), which together form ROC-COR domain, followed by the kinase domain [134, 149]. These domains are flanked by the N-terminal armadillo (ARM), ankyrin (ANK), LRR and c-terminal WD40 domains, which are regulatory protein-protein interaction domains [149].

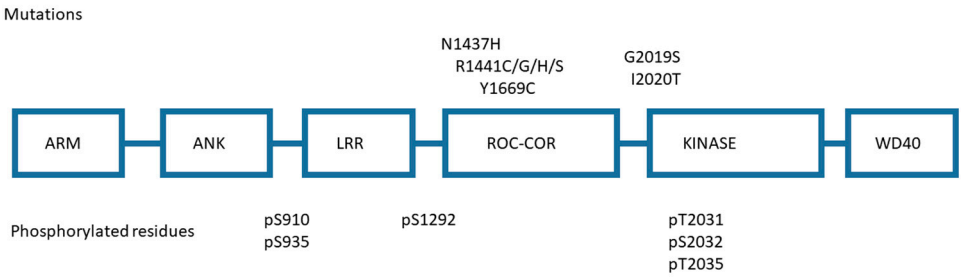


Figure 2 Schematic overview of LRRK2 domains, common mutations and relevant phosphorylated residues discussed in this thesis.

3.2.1 LRRK2 Kinase domain

Two of the LRRK2 PD associated mutations are in the LRRK2 kinase domain: glycine 2019 to serine (G2019S) and isoleucine 2020 to threonine (I2020T), both of which are located at the activation segment of LRRK2 (Figure 2.), which spans amino acids 2017-2054. Both mutations are located at a conserved kinase DFGψ motif (Figure 2.), which in the active DFGψ -in conformation binds to a magnesium ion necessary for catalytic activity. Instead of the conserved DFGψ, LRRK2 has an unusual DYGI motif, where the conserved Phe residues is replaced by Tyr and serves to stabilize the kinase in an inactive DFGψ-out conformation, which decreases the catalytic activity [37]. Mutating the tyrosine 2018 residue to phenylalanine (Y2018F) creates a hyperactive kinase activity similar to I2020T and G2019S mutations, although the I2020T LRRK2 activity readout is dependent on the assay being used [37], which results in contradictory results for I2020T activity.

While these mutations lead to hyperactive LRRK2, the mechanisms for causing this are different. LRRK2 exhibits higher kinase activity as a dimer [150, 151], and interestingly I2020T, but not G2019S homodimerize more strongly than wild type [152]. LRRK2 with I2020T mutation spontaneously forms filamentous structures due to association with microtubules [37, 153], which is also seen with the Y2018F mutation [37]. Wild type and G2019S LRRK2 require inhibition by a type I kinase inhibitor for the same effect [37, 154], which effectively stabilizes the kinase in an active closed conformation. The reason for this difference possibly relates to the kind of conformation the kinase adopts. A recent structural study with full length LRRK2 revealed that the inactive wild type (WT) and G2019S have similar structure [155], which along with the cited evidence above suggests a conformational difference between the different LRRK2 mutation forms. Conformation equilibrium for I2020T and Y2018F shifts towards active closed conformation [156], while WT and G2019S LRRK2 form filamentous structures, when this conformation is stabilized with type I inhibitor [37]. Furthermore, type II

inhibition, which stabilizes the inactive DFGψ-out conformation, does not lead to the filament formation and MT association of LRRK2 G2019S [156]. This conformation-based difference can be one of the explanations why I2020T and G2019S mutations both cause hyperactive kinase, but with differing assay results. An additive explanation comes also from structure, as I2020T has a conformational shift towards active closed conformation, which can lead to slower substrate release [157]. In conclusion, G2019S possibly enhances kinase catalytic activity because of altered kinase kinetics, and I2020T through structural change [140, 156].

Detecting LRRK2 kinase activity

The protein kinase activation segment often contains one or more Ser/Thr/Tyr auto-phosphorylation sites 5-10 amino acids N-terminal to the conserved the APE sequence in the activation loop, which become phosphorylated upon kinase activation. There is evidence that LRRK2 auto-phosphorylates itself on Thr2035 [148, 158-160], Thr2031 [159, 160] and Ser2032 [148, 159, 160]. While Thr2035 originally was found be indispensable for kinase activity [158], the opposite result has been reported [148]. An unambiguous auto-phosphorylation site in the activation site indicative of LRRK2 activity has not been established, and several phosphorylation sites located outside the kinase domain are proposed to be indicative markers of activated kinase: Serine Ser910 and Ser935 [161], and Ser1292 [162].

Auto-phosphorylation at Ser1292 decreases with kinase inhibition [162-166] and many pathogenic mutations including I2020T and G2019S show elevated phosphorylation at this site relative to wild type [162]. This site also has an increased phosphorylation in G2019S mutation PD carriers [167] and sporadic PD cases [143, 144], making this site an attractive target in monitoring LRRK2 activity in vivo and biochemical assays.

Two other proposed surrogates for kinase activity Ser910 and Ser935 were discovered to have decreased phosphorylation in presence of various PD associated mutations including I2020T, and act as binding sites for 14-3-3 proteins [168]. Despite the having decreased phosphorylation in PD relevant mutations including in I2020T, these sites respond to kinase inhibition by loss of phosphorylation [161, 164, 169-176], making these sites valuable tools in monitoring pharmacological inhibition. Confounding the matter Ser935 phosphorylation is increased in patient fibroblasts [145], decreased in PD patient brain [177], unchanged in sporadic PD patient peripheral blood mononuclear cells (PBMC) [178] and neutrophils [176] and decreased in G2019S positive PD patient PBMCs [179]. At the same time, mutating these sites to alanine does not affect kinase activity [163, 168]. Furthermore, there is an inverse relationship between LRRK2 Ser1292 and Ser935 phosphorylation levels [144]. Taken together phosphorylation Ser910 and Ser935 are useful for monitoring kinase inhibition efficiency, but using these sites as a

readout of activated kinase is complicated, and phosphorylation at Ser1929 is a more direct readout of kinase activity.

Finally LRRK2 substrate Ras-associated binding (Rab) family member Rab10 [175], Moesin, synthetic peptides [180] and myelin basic protein [158] are used to detect LRRK2 activity. Of particular interest is Rab10, which is a physiological substrate of LRRK2. Rab10 was originally identified as LRRK2 substrate in an unbiased LRRK2 substrate screen [175] and it's phosphorylation at Thr73 is used as a surrogate marker of an activated kinase and kinase inhibition [165, 176, 181, 182]. Phosphorylation state of Rab10 correlates with kinase inhibition and subsequent dephosphorylation of Ser910/935 [165, 175, 176, 183] and Ser1292 [165]. However this phosphorylation did not differentiate G2019S [182] and sporadic patient neutrophils from the healthy [182, 184]. In an urine study Rab10 phosphorylation is unable to differentiate G2019S PD and healthy individuals, while sporadic PD has a minimal difference to healthy [185]. To further complicate the results a recent study using PBMC was able to demonstrate increased phospho Rab10 levels in both sporadic and G2019S PD [186]. Taken together monitoring LRRK2 kinase activity ideally uses multiple readouts as demonstrated by [37]. These readouts are designed for the experiment at hand, and LRRK2 readout can differ between experiments and patient cohorts.

3.2.2 LRRK2 ROC-CORD domain and N- and C-terminal interaction domains

ROC-COR domain is the second catalytic domain in LRRK2 located N-terminal to the kinase domain and can activate the kinase domain in a GTP bound state [158, 187-189]. Mutations causing amino acid substitutions in three residues in this domain are associated with PD: asparagine 1437 to histidine (N1437H), arginine 1441 to cysteine/glycine/histidine (R1441C/G/H) and tyrosine 1669 to cysteine (Y1699C) [190]. ROC-COR domain functions as a GTPase domain and the mutations in this domain reduce GTPase activity and ROC-COR function [158, 187-189, 191]. This leads to increased kinase activity against non-physiological substrates [158, 187-189], and the physiological substrate Rab10 alike [175, 182, 192, 193].

Mechanistically mutations at R1441, N1437H and Y1699C prolong the GTP bound state by disturbing proper ROC domain dimerization and therefore decreasing GTPase activity [194-196]. Since LRRK2 kinase activity requires GTP [158] it is conceivable, that attenuated GTPase activity leads to prolonged LRRK2 signalling in the subcellular localizations where LRRK2 is recruited to.

N- and C-terminal interaction domains contain less common disease-causing mutations. The N-terminal domain consists of armadillo, ankyrin and leucine-repeat rich domains, while the C-terminal regulatory domain is WD40. These domains are

largely regulatory in function, acting through protein-protein interactions. Removing the N-terminal regulatory domains creates constitutively active LRRK2 kinase, possibly through the relief of WD40 interaction with N-terminal domain [156]. WD40 domain appears to be involved in dimerization and subsequent neurotoxicity of LRRK2 [197], as well as involved in interaction with vesicles [198] and the microtubule system [153, 198]. The best-characterized regulatory process through interaction domains occurs through armadillo and ankyrin. Here Rab29 recruits LRRK2 to membranes, and activates LRKR2 kinase activity through interactions with ankyrin [199], and armadillo domains [200]. Additionally, Rab10 and Rab8 are able to bind armadillo domain, where Rab8 binding stimulates phosphorylation of Rab10 [200]. In addition N-terminal domains interact with FAS-associated death domain protein [201] and E3 ubiquitin ligase CHIP [202].

3.3 Molecular mechanisms of PD

Parkinson's disease is a diverse disease, with several genes involved in multiple cellular pathways. The first PD associated mutations were discovered in the gene coding for αSyn in the late 1990s in unrelated Greek families [203]. Since then multiple genes have been associated with PD or are putative risk factors [132], and although not all of their functions are not clear, one way to categorize PD mechanisms is to look at broad categorical functions, in order to contextualize the complex biology. In this thesis, I will breakdown these functions into mitochondrial dysfunction, disturbed proteostasis and inflammation.

3.3.1 Mitochondrial dysfunction

The brain consumes ~20% of the body's glucose [204], which puts a large demand for brain mitochondrial function, which is a function that has been shown to be impaired in both sporadic and genetic forms of PD. Exposure to mitochondrial complex I inhibitors 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) [205] and rotenone [206] both cause PD symptoms. At the same time decreased complex I activity is observed in sporadic PD patient brain tissue [207, 208] and other tissues such as muscle [209]. It is possible that in sporadic PD the loss of complex I activity is related to damage caused by reactive oxygen species (ROS) [208], which is consistent with signs of oxidative stress in PD patient substantia nigra [210]. As mitochondria are a major source of intracellular ROS, it is possible that in sporadic patients the oxidative damage seen in the brain is caused partly by dysfunctional mitochondria. Alternatively, the mitochondria might be damaged by ROS generated by dopamine metabolism. SNpc dopaminergic neurons at basal levels contain high concentrations of ROS due to dopamine metabolism in neurites [211]. As axon retraction in PD appears to precede dopaminergic neuron death, one possibility is

that either ROS overload at dopaminergic neuron pre-synapses, from mitochondria or dopamine metabolism is one of the causes behind axon degeneration.

Several PD relevant genes such as LRRK2, DJ-1 and SNCA are linked to mitochondrial dysfunction. LRRK2 G2019S mutation carrier patients have decreased mitochondrial membrane potential and intracellular ATP levels indicative of mitochondrial dysfunction [212]. Additionally, αSyn has been shown to co-localize with mitochondria in rodent neurons [213], and appears to sequester mitochondria in LB formation [214] and several SNCA mutations are associated with mitochondrial damage [215]. Protein/nucleic acid deglycase DJ-1 binds directly to ATP synthase complex, and loss of DJ-1 leads to mitochondrial depolarization [216]. The most studied genes in mitochondrial pathology are PINK1 and E3 ubiquitin ligase 3, which will be discussed below more in detail.

Autosomal recessive loss of function mutations in PINK1 [217] and Parkin [218] are involved in early onset recessive PD. The earliest evidence showed impaired mitochondrial function, accumulation of ROS and loss of mitochondrial proteins by mass spectrometry in Parkin knockout mice brain [219]. Later drosophila models show mitochondrial dysfunction in PINK1 knockout flies, and genetically linked PINK1 and Parkin to same pathway [220, 221]. This pathway activates when mitochondrial function is impaired, and starts with PINK1 recruitment to outer mitochondrial membrane, where PINK1 recruits Parkin [222] by phosphorylating Parkin at its ubiquitin like domain [223-225], and Ubiquitin [21, 224-227]. This in turn results in ubiquitination of mitochondrial Parkin client proteins, and clearance of damaged mitochondria in a process called mitophagy [228]. One of the hallmarks of damaged mitochondria is an increase of intracellular reactive oxygen species. Interestingly PINK1 activity is negatively regulated by reactive oxygen species, which raises the question of PINK1 dependent mitochondrial mechanisms [229]. Additionally it appears that mitochondrial basal mitophagy in mammalian brain occurs independent of PINK1 [230], and more studies are needed to understand processes that PINK1 and Parkin regulate.

As rotenone directly causes PD, LRRK2 is implicated in mitochondrial dysfunction, and mitochondrial dysfunction is observed in PD we chose rotenone to model PD in vitro and rodent rotenone model of PD (Paper I).

3.3.2 Proteostasis

Proteasome function, αSyn *and Lewy bodies*

The proteasome is a large protein complex consisting of a catalytic core with trypsin and chymotrypsin like activity, and multiple regulatory and structural subunits and is responsible for degradation of cytosolic and nuclear proteins tagged with ubiquitin [231]. Proteasomal function has been observed to be impaired in PD [232, 233] and mutations in proteasome subunit S6 ATPase are associated with early onset

form of PD [234]. One of the most prominent histological signs of PD in many, but not all cases is the formation of LB. The major component of LBs and driver of their formation is thought to be αSyn, which is found to be ubiquitinated in cells, indicative of a malfunctioning ubiquitin-proteasome system [235].

Under basal conditions, αSyn is localized to presynaptic membrane structures [236-238], and is thought to regulate synaptic transmission. In simplified terms in the non-pathogenic state, αSyn exists as a monomer but due to some unknown mechanism forms various fibril polymers, which over time turn to form LB inclusions. Studies show that αSyn exhibits prion like properties [239], and the prion propagation can start from the peripheral tissues such as gut [240, 241]. In connection to this artificially preformed synuclein fibrils are able to convert soluble αSyn to aggregates [242, 243], and interestingly αSyn derived from patient LB are able to seed inclusions that resemble LB found in the disease in terms of abundance and subcellular distribution [244]. Furthermore, a recent study using an in vitro test demonstrates that formation of these inclusions by preformed inclusions is highly specific to individuals suffering from PD compared to healthy, non-manifesting mutation carriers and motor disease patients without dopamine deficit [245].

Duplication and triplication mutations in αSyn are linked to PD [132] suggesting a gene dose dependent effect. Despite this, it is not fully clear if formation of LB is one of the drivers for pathogenesis or is symptom of the disease [246]. There is evidence that functional αSyn is required for normal neuronal function, as knockdown in non-human primates [247] and in the rat [248] causes dopaminergic neuron degeneration, and neurological impairment and premature death in mice [249]. Around 90% of the αSyn found in LB is phosphorylated at Ser129 [250, 251] and it is unclear if phosphorylation at this residue either precedes or is downstream of the LB formation. Contrary to this phosphorylation event being disease causing, a recent study demonstrates that this phosphorylation correlates with normal neuronal activity [252], which makes the consequence of Ser129 phosphorylation unclear and further studies are necessary. In summary, it is clear that one of the major genes: αSyn is involved in disturbed proteasomal function, which as a process is a disease-causing mechanism in PD, but there are many unanswered questions.

Autophagosome and lysosomal function

Autophagy is a process where cells degrade cellular material, by either engulfing the cargo in a double membrane vesicle and delivering the contents to lysosome, or directly delivering the cargo to lysosome [253]. As mentioned earlier mitophagy, which is a special form of autophagy, is involved in PD pathogenesis. This form of organelle specific autophagy is referred as selective macroautophagy in the literature, which targets organelles such as ER or misfolded protein aggregates. A second kind of macroautophagy: non-selective autophagy targets bulk cytosol to degradation to provide nutrients during starvation [253]. In both cases, a double envelope forms around the bulk cytosol or organelle, which then fuses with

lysosome to form autolysosome, where hydrolases degrade the contents. Interestingly autophagy-lysosomal function is disturbed in both familial and sporadic Parkinson's disease. Markers indicative of increased autophagosome formation are present in PD post-mortem brains [254]. At the same time, lysosomal markers are decreased in PD patient brain [255-258]. Taken together accumulation of autophagosomes appears to be a consequence of impaired lysosomal function in both rodent PD models and in PD patient SNpc neurons [257]. Furthermore, accumulation of autophagic markers is evident in LB, which again is indicative of impaired lysosomal function [257].

Several genes mutated in PD have a strong links to autophagy functions. Most notably lysosomal enzyme Glucocerebrosidase (GBA) gene contains multiple loss of function mutations associated with PD causing lysosomal dysfunction [259]. This enzyme functions to cleave glycosphingolipids into their constituent parts, and sporadic PD patients have decreased GBA activity [260, 261]. Even though the activity decreases in PD, healthy individuals also display decreasing GBA activity over the course of normal aging, eventually reaching similar activity levels as in PD [260], which suggests further mechanisms that have not been observed so far. Interestingly patients harbouring both GBA and LRRK2 mutations have less severe disease phenotype than patients harbouring just the GBA mutation [262, 263], suggesting that LRRK2 and GBA converge on same pathway, through an unknown mechanism.

As mentioned earlier several Rab proteins are LRRK2 substrates [165, 175]. Rab proteins are GTPases, which are kept inactive in the guanosine diphosphate (GDP) bound mode in cytosol, and in the active GTP bound form they associate with membranes, and facilitate vesicular transportation [264]. Once a Rab cargo has reached the target membrane, Rab GTPase activity is activated, and Rab is stabilised in the GDP bound mode by their regulatory proteins, and recycled back to proximity of their original membrane [264]. A consequence of the LRRK2 dependent Rab phosphorylation appears to be loss of Rab interacting protein interactions trapping Rab at their target membrane interfering with normal membrane trafficking of Rab proteins [175, 265]. Not surprisingly, LRRK2 mutations lead to endo-lysosomal dysfunction in rodent and drosophila models [266] and this dysfunction is evident in sporadic PD brain neuron soma [258]. Furthermore, rotenone treatment, which is a known activator of LRRK2, leads to reduction in GBA activity and end-lysosomal dysfunction in neuronal soma, all of which are reversed by LRRK2 inhibition [258]. Specific to neurons, LRRK2 kinase activity leads to impaired retrograde autophagosome transportation in neurons in both G2019S and R1441H models [267, 268], which is done through phospho-Rab dependent activation of kinesin [268]. Taken together suggests that both lysomal maturation and function, as well as transportation of autophagosomes are impaired in PD.

Protein synthesis and unfolded protein response

Protein synthesis is a tightly regulated process involving multiple regulatory processes. One of these processes is the unfolded protein response (UPR), which activates when unfolded proteins accumulate in endoplasmic reticulum. This process serves to prevent further accumulation of unfolded proteins by slowing down protein synthesis and directing cells to make chaperones. Sustained activation results in activation of pro-apoptotic JNK signalling cascade and caspases. Here type I transmembrane protein inositol requiring 1 (IRE1) and activation transcription factor 6 (ATF6) pathways regulate the expression of proteins involved in UPR. A third pathway of UPR is protein kinase RNA-like endoplasmic reticulum kinase (PERK), which upon activation phosphorylates eukaryotic initiation factor 2 alpha (eIF2α), resulting in arrest of cell cycle and protein synthesis [269]. Under resting conditions, 78-kDa glucose regulated protein (GRP78) binds to IRE1, ATF6 and PERK and keeps them inactive. GRP78 preferentially associates with unfolded proteins, and upon accumulation of unfolded proteins, GRP78 dissociates from these three proteins resulting in activation of ATF6, IRE1 and PERK [269].

Relevant to PD activation markers of PERK were found to be present in PD patient brains [270-272], while the ER chaperone GRP78 is decreased [272], indicating dysregulated UPR and a decreased capacity for unfolded proteins. Furthermore, a member of protein-disulphide isomerase (PDI) family proteins, which help to mature unfolded secretory protein levels, is increased MPTP model and PD patient brain, and is present in LB [273], which could indicate a compensatory mechanism in response to unfolded proteins. S-nitrosylation of PDI family proteins can cause their inactivation, which contributes to endoplasmic reticulum (ER) stress and UPR activation, and this modification is increased in PD brain [274]. This modification is also induced by rotenone exposure [274], which suggests that mitochondrial dysfunction and increased ROS levels could activate UPR in PD. Rotenone appears to affect other components of UPR pathway, as it is able to induce phosphorylation of eIF2α in LRRK2 dependent manner in cultured neurons [145] (Paper I). Additionally many components of UPR are activated in response to rotenone treatment in an αSyn phospho-Ser129 dependent manner in a αSyn expressing cell line [275]. Further evidence of LRRK2 and αSyn involvement in UPR comes from a study where LRRK2 interacts with several ER resident proteins and regulates ER Ca^{2+} levels [276]. The treatment of cells with αSyn appears to trigger calcium depletion in the ER, which is exacerbated in cells derived from transgenic G2019S mice, leading to activation of UPR [276]. Taken together these results, and many more that are not cited in this thesis, suggests that PD mutations and environmental toxins predispose individuals to PD and UPR is involved.

Connected to UPR, aberrant protein synthesis is present in PD through other mechanisms. The first line of genetic evidence comes from a family in Northern France with an autosomal dominant form of late PD, where eukaryotic initiation factor 4-gamma 1 was found to contain PD associated mutations [277]. Increased

elongation factor 2 phosphorylation (EEF2) indicative of decreased protein synthesis is present in PD brain [278] and in an in vivo rodent rotenone model of PD [145] (Paper I). Fibroblasts derived from sporadic and LRRK2 G2019S PD patients also show increased EEF2 phosphorylation in addition reduced protein synthesis [145] (Paper I). Proteomic study of the same cells revealed that proteins related to proteostasis and protein synthesis are down regulated relative to healthy controls [279] (Paper II). Contrary to evidence from patient derived samples, LRRK2 G2019S over-expression models in drosophila and cultured human neurons showed an increased protein synthesis mediated by ribosomal protein s15 phosphorylation [280]. Pharmacological inhibition of endogenous LRRK2 protein levels rescues reduced protein synthesis in sporadic and G2019S patient fibroblast under basal conditions, and in cultured neurons in the rotenone model [145] (Paper I). Different results obtained by Martin et al. [280] possibly relates to differences between overexpression and endogenous LRRK2 levels found in cultured neurons and patient fibroblasts, and an open question is, which direction protein synthesis changes in cells with high and low LRRK2 levels. Ribosomal profiling of LRRK2 WT and G2019S knockout brains show equal differences in both directions [281], which indicates that the issue is complex, although ribosomal profiling does not take translational efficiency into account. Interestingly significantly different protein levels for ribosomal protein in both cytosol and mitochondria showed almost unilateral downward trend in PD patient brain, in a recent proteomic study [282], which again would suggest impaired protein synthesis in neurons. Taken together it appears like that protein synthesis machinery and its key regulators are dysregulated in PD.

3.2.3 Inflammation

Inflammation in PD was first described by McGee and colleagues in 1998 with the observation that microglia were activated in PD patient sNPC microglia [283]. Additionally CD+4 positive lymphocyte invasion to PD brain has been observed [284], and CD14+CD16- monocytes have been observed to increase in PD patient blood [285]. Conversely CD+4 lymphocytes are decreased in the periphery in multiple studies, while other cell types show mixed results [286]. Additionally multiple studies show increased neutrophil count in PD patient blood [287]. Recent study replicates decreased lymphocyte count with two large independent patient cohorts, and also shows a trend for increased neutrophil count in sporadic and GBA PD, but not in LRRK2 PD [288], which suggests different inflammatory mechanisms for different types of PD.

In addition to cellular mediated immunity, the complement system has been shown to be activated in PD sNPC Lewy Bodies [289, 290]. Furthermore, complement protein levels increase in CSF in multiple studies [291]. Not surprising complement components have come up as top biomarkers in recent proteomics high depth studies

from patient samples coming from various sample materials sources including brain [282, 292, 293]. Additional evidence of complement activation comes from a cellular model, where overexpression of αSyn activates the classical complement pathway [294].

The most prominent gene with mutations in PD with strong links to inflammation is LRKR2. LRRK2 is highly expressed in various immune cells [146, 295, 296], especially in neutrophils [176]. In addition, LRRK2 is involved in inflammatory bowel diseases (IBD) [139], which is a risk factor for PD. A Taiwanese population study showed that individuals suffering from IBD had 35% higher risk of developing PD [297], which yet again suggests that PD starts from the periphery. In disease models, lipopolysaccharide (LPS) treatment in mice causes delayed loss of dopaminergic neurons in mice by 7 months after systemic delivery of LPS [298]. This is not observed in transgenic wild type LRRK2 mice in response to LPS treatment, while LRRK2 R1441G or G2019S mutation mice have loss of dopaminergic neurons in SNpc [299]. Interestingly many immune cell types have higher LRRK2 protein levels in sporadic PD relative to healthy controls [295, 296] and many cytokines are able to elevate LRRK2 protein levels in cultured PBMC [300]. Pathogenic mutations in LRRK2 can increase inflammatory cytokine and chemokine levels, in kinase dependent manner [301]. Other Parkinson's disease related genes also link to PD, and functionally most likely alter the immune system in different ways [286]. Taken together multiple lines of evidence demonstrate that immune system is disturbed in both sporadic and genetic PD alike.

4. Proteomics

Advances in the study of polynucleotide deoxyribonucleic acid (DNA) analysis techniques have allowed the development of global gene wide association studies (GWAS) [302], with the aim of understanding how genomic variance explains physical phenomena such as physical and mental traits, as well as disease states. Although providing valuable information, these studies fail to fully explain the resultant phenotype traits such as body weight, intelligence or personal preference between coffee and tea. GWAS studies also fail to fully explain when or if ever, an individual carrying disease associated variant(s) will develop the disease.

Another polynucleotide-based approach: transcriptomics measures ribonucleic acid (RNA) levels, tries to bridge the gap between DNA and the phenotype, and is popular due to mature sample preparation and measuring methods. Since messenger ribonucleic acid (mRNA) is the blueprint for making proteins by ribosome, the assumption is that mRNA levels reflect protein levels and the biological complexity. Proteins in turn largely make up the biological state of a cell or tissue and ultimately a phenotype. However, merely looking at mRNA levels alone is not sufficient to explain function of the protein present in the cell. Proteins contain PTMs, have dynamic structures, they interact with each other and metabolites in various complexes and subcellular locations. In addition, proteins have varying states of solubility and have variable half-lives, information that is invisible at mRNA level. Furthermore, diagnostically valuable body fluids such as plasma or urine have no active translation, and a phenotypic readout on changes occurring at protein level is only possible by direct protein readout. Below I will be discussing how Mass spectrometry-based proteomics works from sample preparation to Mass spectrometric analysis methods. Finally, I will be discussing how making protein level correlations from mRNA under steady and non-steady state works, and briefly introduce non-Mass spectrometric proteomic techniques. Various RNA species also have other important functions, such as regulating transcription, transcript slicing, protein synthesis and protein activity [303], but will not be discussed in detail in this thesis.

4.1 Mass spectrometry based proteomics.

As discussed above, the prerequisite for understanding any phenotype requires the study of the proteome. Traditionally this has been achieved with classical biochemical techniques such as Edman degradation [304] or probe-based technologies such as Western blotting and it's derivate eastern blotting [47, 305]. However, these techniques are not high throughput and do not provide complete protein coverage and can yield unspecific signals. In answer to this Mass spectrometry based proteomics, is the key technology that has enabled the study of proteins at high throughput at the proteome level, with a diverse array of instruments, techniques, and applications.

The current estimate is that that each cell type expresses roughly 10k genes at a time, with a detectable protein product [306]. Each gene can produce multiple transcripts through splicing, each transcript can produce multiple forms of proteins through alternative start codons and may contain individual genetic variations. Proteins in turn are post-translationally modified, and each protein with distinct set of variables coming from a gene is called a proteoform [307] (Figure 3A.). It is not clear how many proteoforms are present in any human cell type, but Aebersold et al. estimated one million [308], which is a striking number considering humans have roughly 20k genes. Currently mass spectrometry-based proteomics is the most mature technique used to characterize quantitative protein level differences between conditions, PTMs, protein turnover, structure and various proteoforms that make up the proteome, and in the end finding the determinants that define a phenotype.

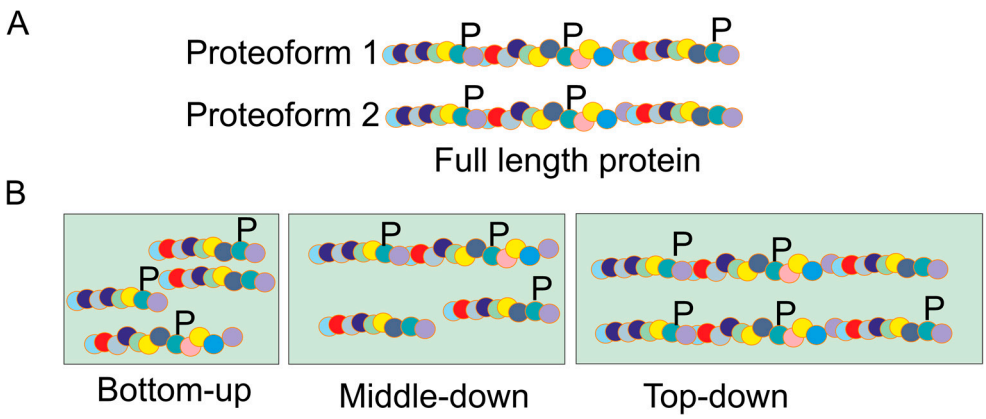


Figure 3. Multiple proteoforms with distinct set of modifications from the same gen can exist, such as different sites being phosphorylated in a proteoform (P on the protein chains) A. Bottom-up proteomics measures cleaved proteins, making it challenging to assign peptides to particular proteoform, while middle-down and top-down proteomics allow for better sequence coverage and proteoform assignment B.

Mass spectrometers measure charged ions by their mass to charge ratio (m/z) in either intact ion (MS1), or in fragmentation mode (MSn). Proteomic samples are very complex and are usually separated by chromatography to allow for a more in depth analysis. Intact analytes in proteomic samples often have an identical m/z, and to identify the analyte, its peptide/protein sequence must be established. This is achieved by MS2 analysis, where the intact precursor ions are first detected by MS1 scan, and then fragmented inside the Mass spectrometer. The resulting MS2 spectrum provides the sequence information necessary for establishing identity of the intact ions. If necessary, further rounds of MSn analysis can be performed to obtain unambiguous identification.

A Mass spectrometric analysis of proteome has the following stages: sample preparation, separation, ionization, mass analysis, optional fragmentation and ion detection and data processing and analysis. The actual way these are done varies and the choice starts from what is being measured. The proteome can be investigated by Mass spectrometry in three ways: bottom up, middle down, top down (Figure 3B.), each of which have multiple analysis strategies. Bottom-up proteomics typically measures ~6-30 amino acid length peptides generated from a protein sample by enzymatic digestion, middle down proteomics peptides larger than 30 amino acids derived from chemical or enzymatic digestion and top-down intact proteins. Top-down analysis provides the most information rich detail on a proteoform, but it is, as of writing limited to detecting several thousand proteoforms coming from several hundred proteins [309]. This is largely due to challenging sample preparation, the large dynamic range of the proteome and difficulties in separating the analytes necessitated by complexity of the proteome, and fragmenting the proteins in MS2 analysis and the subsequent data analysis [309, 310]. Novel separation methods such as circular ion mobility and proton transfer charge reduction are being developed and may eventually solve these problems ([311, 312]. Middle-down is a compromise between bottom up and middle down, but largely suffers from the same problems as top down. Additionally, there are no standardized and easy to use sample preparation methods available for generating large peptides and analysing them reliably [313].

Bottom-up proteomics is the most common proteomic sample preparation strategy in use, largely owing to its flexibility and relatively easy sample preparation. Another advantage of bottom-up proteomics is that modern Mass spectrometers can resolve and fragment small peptides more readily, and the identification and the analysis software is mature. The largest disadvantage of bottom-up proteomics compared to middle and top down is limited sequence coverage of the protein, which leads to ambiguous protein identity assignment, as many proteins from closely related families or functions contain identical protein sequence segments. When such segments are detected, it is impossible to tell unambiguously, which protein they belong to. The second disadvantage relates to post-translational modifications. Proteins contain multiple post-translational modifications and relationship between

different post-translational modifications and the associated proteoform is lost in bottom-up proteomics, as it is impractical to design experiments to detect multiple PTMs at a large scale. Despite these disadvantages related to bottom-up proteomics, the techniques associated with it are most the most matured in proteomics, and were used in the papers covered in this thesis (Paper I-III).

Modern Mass spectrometers are limited in their detection range, and typically can detect protein quantities in the range of 3-7 orders of magnitude in a single shot experiment. This depends on the instrument analysis mode used, where older Orbitrap instruments have dynamic range of over 3 orders of magnitude [314], and ~5 orders of magnitude with the newer Orbitrap [315, 316] and time of flight [317] instruments in untargeted mode. The cellular proteome has dynamic range of roughly 7 [318] orders of magnitude and bio fluids such as plasma 12 orders of magnitude [319]. Therefore, to fully understand protein level differences, post-translational modifications and how they translate into a particular phenotype, special fractionation and enrichment techniques are commonly used to dissect proteome level changes occurring outside the detection range. The scope of this thesis lies in the domain of bottom-up proteomics and following chapters will be used to discuss steps involved from sample preparation to ion detection of bottom-up protein sample. In the remaining part of the thesis, I will discuss sample preparation methods for bottom-up proteomics, followed by Mass spectrometry data acquisition modes, data processing and quantitative analysis of proteomic MS data, and finally non mass spectrometry based methods to study the proteome.

4.1.1 Sample preparation for bottom-up proteomics.

The efficient analysis of bottom-up samples by Mass spectrometry requires that the peptides, which are used in the analysis, are ~7-30 amino acids in length. The most common way of achieving this is to enzymatically digest the sample with trypsin, which specifically cuts C-terminal to arginine and lysine, which usually generates > +2 charged ions, and are highly abundant on protein surfaces, and generate peptides that can be used to detect most of the proteome [320, 321]. There are other enzymes used for bottom-up proteomics such as chymotrypsin, which cuts preferentially at aromatic amino acid residues and at lesser rate at leucine and methionine, which is advantageous when analysing highly hydrophobic proteins such as multi-pass transmembrane proteins.

Digestion for bottom-up proteomics.

There are multiple ways to achieve digestion of protein sample, and these can be largely divided in to three categories: in solution, gel-based digestion methods and device-based methods, which have been extensively compared [322-326]. The choice of which method largely depends on combination of scale of the experiment, cost, reliability, and the downstream application.

In solution digestion is a commonly used approach in bottom-up proteomics. Here the sample is efficiently denatured followed by reduction and alkylation of cysteine side chains, followed by trypsinization and sample clean up. Samples for in solution digestion samples are often denatured with Urea or Guanidine hydrochloride, both of which are chaotropes. Ionic detergents such as sodium dodecyl sulphate (SDS) sodium deoxycholate (SDC), the alcohol trimethyl,2,2,2-trifluoroethanol and the commercial MS compatible detergent RapiGest™ are also extensively used [327-329]. Samples may contain MS incompatible components such as detergents, lipids, and metabolites, which can be removed by protein precipitation if needed [322]. The downside of chaotropic denaturants is that, even when they are diluted, they interfere with the activity of the trypsin, leading to more incomplete digestion than in other methods [322, 323, 327]. Another downside of in solution methods is that the sample preparation and digestion introduce MS incompatible chemicals, which must be removed prior to MS analysis. Salts, water soluble chemicals and chaotropic reagents are effectively removed by reverse phase (RP) C18 clean up. Ionic detergents SDS [328] and SDC [329] can be removed by precipitation, which is often followed by RP C18 clean up to further clean and concentrate the samples. A recently developed in solution method SPEED uses trifluoroacetic acid for solubilisation. SPEED samples are cleaned by RP C18 after reduction, alkylation and trypsinization have been done in the same sample tube, which streamlines the sample handling [330]. This method is simpler than the older in solution methods and compares favourably to modern digestion methods in terms of reproducibility and data depth [322]. While these methods ensure that the sample is MS ready, RP C18 clean up adds an extra cost and handling, and roughly 50% of the sample is lost during the clean-up. The advantages of in solution methods are the relatively cheap cost and simplicity in sample handling.

In the gel-based methods, the sample is either embedded or separated in a polyacrylamide gel under native or denaturing conditions. Following this, samples are stained to allow visualisation for subsequent excision from the gel. In the next step, the gel diced into small pieces, destained, and in the process cleaned thoroughly from any MS containing contaminants in the process. Following destaining, cysteines are reduced and alkylated, and trypsin is added to gel pieces [331]. The sample obtained after peptide extraction is ready for downstream experiments or liquid chromatography Mass spectrometry (LC-MS) analysis. In gel digestions are labour intensive and slow, but show robust results [324], although the peptide yield is strongly affected by input amount and is smaller than with other methods under certain conditions [325, 326]. In gel digestion is possible to automate, but is still limited by the necessity of running SDS-PAGE gels and the manual excision of the sample lanes [332], which limits the use in large scale experiments. The advantage of in gel digestion is that samples are cleaned thoroughly through the handling, and it is possible to fractionate proteins of interest by size, which allows for characterization of proteins of interest and proteoforms based on size and cross-linked protein complexes.

Device based methods are a more recent development, relatively speaking. Protein samples are generally denatured with strong denaturants, followed by reduction and alkylation. The samples are trapped physically on or by a solid matrix during or after the denaturation and alkylation. Trapped samples are extensively washed on the devices effectively removing most interfering substances, which increases efficiency of both digestion and streamlines downstream handling. The advantages of trapping methods is their speed, and they often give pure peptide samples, which are ready for downstream analysis. There are many commercial options such as S-Trap [333], SP3 [334], Si-Trap [335] and iST [336], and in house methods such as FASP [337]. Many of these methods are available in 96-well plate format, and owing to their simplicity, they are increasingly used in high-throughput studies. SP3, S-Trap and Si-Trap generate samples that are ready for simple protein level analysis without the need of sample purification, while FASP and iST require this. FASP is the non-commercialized method and is cheaper than the commercial methods, but the method involves multiple long centrifugation steps, which makes it more time consuming. The commercial device methods discussed are more streamlined, but at the same time are more expensive than FASP or the in solution and in gel methods.

Bottom-up approaches for newly synthesized proteomes.

As total protein levels are made up of different proteoforms in various states of activity, solubility and processing, it is beneficial to study newly synthesized proteome, to assess response to drug treatments or perpetuations in protein synthesis. Historically protein synthesis was studied by ribosomal occupancy and S^{35} labelling. However, these techniques do not give global information about which proteins are undergoing synthesis. Non-canonical amino acids probes like L-azidohomoalanine (AHA) [338] and L-homopropargylglycine (HPG) were developed [339] and are incorporated in to growing polypeptide chains in place of methionine. These amino acids contain an azide side chain, that can be conjugated to alkyne containing molecules by click chemistry [340], and proteins tagged with these amino acids can be enriched and analysed by MS using a technique like BONCAT [341]. Here, azide containing proteins are covalently bound to alkyne agarose, allowing for the enrichment, and are then digested on bead, followed by MS analysis. However, as the AHA residues that unambiguously identify newly synthesized proteins are covalently bound to alkyne resin after the digestion, it is difficult to tell newly synthesized protein from proteins that bind agarose by unspecific binding. As further development to BONCAT, Direct Detection of Biotin-containing Tags method first digests the protein, then uses click chemistry to conjugate AHA containing peptides to biotin alkyne, and enrich biotinylated peptides on streptavidin agarose leading to unambiguous identification and increased sensitivity compared to protein enrichment based technique [342], and is applicable to in vivo labelling [343].

Labelling with AHA and HPG is neither fast nor 100% efficient, as cells prefer methionine to analogues by a factor of several hundred times [338, 339], and therefore requires methionine starvation in cell culture or replacement of methionine in the diet in vivo. This puts constraints on short time course experiments and methionine starvation results in a non-native state for the cells. An alternative for non-canonical amino acids, puromycin allows for rabid labelling of newly synthesized polypeptides [344], and an alkyne containing puromycin analogue, can be enriched for MS analysis [345]. Puromycin is incorporated in to growing polypeptide chains in place of tyrosine and upon incorporation, the polypeptide is released from the ribosome [346]. Puromycin containing polypeptide products are cleared away within hours, which creates problems for long time course experiments [347]. Another downside of puromycin-based assays is that it creates proteotoxic stress through accumulation of unfinished protein products that have to be cleared away. Seemingly circumventing the problems with AHA and HPG, which are methionine starvation and long labelling, and proteotoxic stress and protein degradation with puromycin THRONCAT was developed [348]. Here cells are efficiently labelled with the threonine analogue β-ethynylserine in complete media without amino acid starvation. Although β-ethynylserine competes with threonine, efficient labelling is achieved in as little as in an hour in threonine containing complete media. This allows for pulse labelling of proteins in short treatments and appears to be promising tool in detecting changes in newly synthesized proteome.

Newly synthesized proteins can also be studied in a cell type specific manner by targeting transfer RNA (tRNA). This technique utilizes mutant aminoacyl tRNA synthetase, and the methionine analogue L-azidonorleucine (ANL). Here cells expressing mutant aminoacyl tRNA synthetase are able to charge methionine tRNA with ANL, which leads to efficient incorporation of ANL to proteins in place of methionine [349]. This allows for efficient incorporation of ANL in presence of methionine containing medium, in cell type specific manner although this mutated synthetase has 4 fold preference for methionine [349], this is significantly more efficient than AHA and HPG based, which have 100 fold smaller affinity for incorporation compared methionine. At the time of we performed proteome analysis of newly synthesized proteins AHA approach BONCAT [341] was characterized the best, which why we chose this approach (Paper II).

4.1.2 Analysis of bottom-up samples on a mass spectrometer

The choice of the quantitation and data acquisition methods is critical in MS based proteomics experiment, and largely determines quantitative accuracy, proteomic depth, and the cost of analysis. Mass spectrometers can be operated in data dependent acquisition (DDA) mode, where the mass spectrometer first performs an intact ion scan (MS1) (Figure 4A.) followed by MS2 sequencing of most intense

MS1 ions either by number of specified MS2 scans, or under a specific time interval (Figure 4B. and C.). In an alternative data acquisition method, data independent acquisition (DIA) the intact ion mass range is divided into mass windows and windows are sequentially analysed in MS2 mode (Figure 4C.). The third data acquisition mode is targeted mode, where the mass spectrometer is monitoring a specified list of precursor ions. Lastly, a rough cut can be made into unbiased discovery acquisition methods and targeted methods, although with introduction of hybrid DIA method [350], which combines targeted method and unbiased label free DIA, the line between these two is blurring. Although not all encompassing below, I will be discussing the most used data acquisition methods and assessing their strengths and weaknesses.

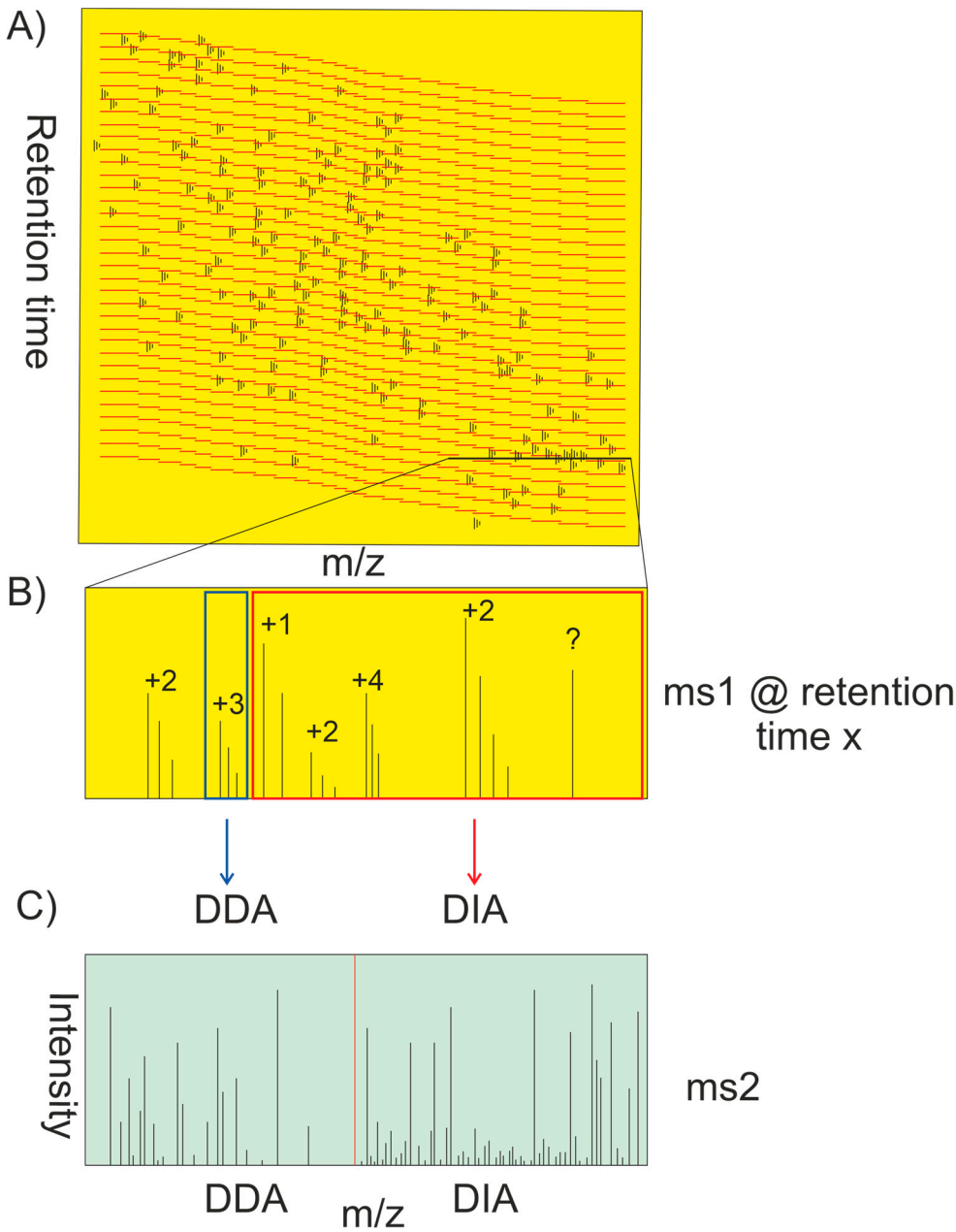


Figure 4. Peptide ions are detected in MS1 scan A. In DDA single peptide ion is chosen for analysis, and in DIA m/z window B. Resulting spectra in DIA analysis are more complex and require more sophisticated analysis software C.

Label free DDA quantification

Label free quantification (LFQ) DDA (LFQ-DDA) quantification typically obtains the sequence information from an MS2 scan and peptide quantity from the precursor ion intensity measured in MS1 scan, as MS2 scans occur randomly outside the MS1 peak apex. This can contain co-eluting peptides and thus does not reproducibly capture peptide quantitative information in DDA [351]. This is relatively inexpensive and quantitatively accurate but suffers from the stochastic nature of DDA precursor ion selection, which increases as a problem as increasing number of runs are being compared. As LFQ-DDA quantification is typically done with MS1 precursor intensities, the problem of missing values with this technique is partially circumvented by matching unidentified MS1 features in one run to an identified MS1 feature with the same mass and retention time in another run as exemplified by MaxQuant [111, 352] also called match between runs (MBR) in the literature. However, despite MBR, label free DDA still suffers from high number of missing values, which is particularly problematic in PTM studies, where a presence of a PTM can be binary, and missing values can cause uncertainty in deciphering the biology. In label free DDA quantification, it is possible to achieve higher proteome coverage by orthogonal fractionation methods with up to 12k protein groups being identified [353-356]. Advanced DDA methods such as Boxcar DDA uses a single shot experiment [357] as does Wide Window Acquisition [358] to achieve the same coverage. All come of these approaches come with some disadvantages. Fractionation in label free DDA comes at the cost of sample throughput, as each sample require multiple fractions in order to achieve high proteome coverage. Boxcar DDA on the other hand heavily relies on assigning unidentified MS1 features to a spectral library without an MS2 scan raising the risk of false positive quantification. Wide Window Acquisition limits the analysis of MS/MS analysis to a vendor specific software (Proteome Discoverer, Thermo Fisher Scientific), and up to now, has not gone through peer review or comprehensive independent evaluation.

Isobaric labelling approach

In an isobaric labelling experiment, the peptide samples are tagged typically with an amine reactive set of labels which have the same nominal mass and appear as single isotopic distribution at MS1 level when the tagged samples are mixed and analysed on mass spectrometer. However, upon fragmentation and MS2 detection each of these labels produce a label specific quantitative reporter ion. There are multiple possible isobaric labelling tags: tandem mass tag (TMT) [359], isobaric tags for relative and absolute quantification (iTRAQ) [360], deuterium isobaric amine reactive tag (DiART)[361], N,N-dimethyl leucine (DiLeu) [362] and 10-plex isobaric tags (IBT) [363]. TMT and iTRAQ are readily available commercial reagents that allow labelling of up to 18 [364] and 8[365], while non-commercially available DiART, IBT and DiLeu reach up to 6, 10 and 21 respectively [361, 363, 366]. Commercial isobaric labels are expensive limiting their use, although efficient labelling can be performed with less reagent than manufacturers recommend [367].

Although there are efforts to combine with targeted methods [368], isobaric labelling experiments are typically done in data dependent mode. Here the precursor peptide chosen for MS2 is a mix of all pooled samples and upon fragmentation and MS2 analysis all samples are simultaneously quantified, which leads to very few missing values within isobaric plex, which is particularly advantageous in PTM studies. As an emerging specialty use, isobaric labelling is also utilized in proteomic single cell methods. Typically eukaryotic cells contain ~100-300 picograms of protein [369], and peptide and protein detection rate rapidly decreases as peptide input closes to pictogram scale even with modern instrumentation [317]. Isobaric labelling circumvents this limitation by utilizing a carrier channel with higher protein amount, pooled with the labelled single cell proteomes, which results in increased MS1 and MS2 signal while maintaining the quantitative information for single cell channels [370, 371].

However, due to stochastic nature of DDA sampling, isobaric labelling experiments can be plagued with missing values between isobaric plexes, with number of missing values reported to reach at 10% and 40% for proteins and peptides respectively at five TMT plexes [372]. Recently an isobaric labelling match between runs algorithm [373] in the popular MS analysis software MaxQuant [111] was developed to alleviate this problem, but missing values between plexes are still restrictive in a large isobaric labelling experiment. Another concern with isobaric labelling is interference signal from co-isolating peptides disturbing the quantitative accuracy, by ratio compression [374-379]. There are several strategies to circumvent this, and in the first, a peptide ion is first fragmented and analysed in by MS2 in ion trap, which is followed by MS3 scan of the most intense product ion in Orbitrap, largely eliminating interference at the cost of number of identified peptides [378]. A second approach is to separate precursor ions in gas phase by mass and charge properties [379] and modern mass spectrometers equipped with ion mobility capabilities are able to take advantage of this idea in isobaric labelling experiments, by reducing the amount of co-isolating ions [380].

The highest proteomic coverage is usually achieved with orthogonal fractionation methods as in LFQ-DDA [355], usually with an off-line high pH reverse phase C18 fractionation followed by online separation of the resulting fractions. Extensive fractionation combined with isobaric labelling not only produce high depth protein level data of up to ~8-12k genes being detected from complex human samples [381-384], and also proteoform specific peptides, abundant PTMs, rare variants and non-canonical proteins are readily detected [381, 383, 385].

Data independent acquisition

In recent years DDA in label free quantification for protein level changes has found a competitor in data independent acquisition (DIA, Figure 4). DIA MS2 spectra are more complex than those acquired with DDA (Figure 4C.), as the product ion species resulting from the fragmentation originate from multiple peptides. This

necessitated building of project specific spectral libraries by deep coverage fractionation, or use of generic libraries, which results in slightly lower quantitative accuracy and identification rates than with project specific libraries [386]. However, library free DIA was enabled by generating pseudo MS2 spectra from DIA scans by matching MS1 peptide features to product ion retention times [387]. This idea has been further developed into multiple DIA analysis pipelines such as MaxQuant [111, 388], DIA-NN [389], Spectronaut [390] and Fragpipe [391, 392], and as of writing this these techniques have come very close to library based DIA in terms of identification. Typically DIA experiments are label free, but recently developed non-isobaric tagging has shown promising results in multiplexing, number of identifications and quantitative accuracy [393]. However, further independent studies are needed to validate this finding.

Targeted methods

Targeted methods fundamentally differ from the shotgun methods described above. Here a list of peptide precursors is monitored either constantly, at certain chromatographic retention time points or the peptide scan is triggered. These methods are typically more sensitive and quantitative accurate than the discovery methods, if designed properly. Method design starts from defining the peptides that are to be quantified. Ideally these peptides do not contain difficult tryptic cut sites, labile or readily modified amino acids such as asparagine and methionine. However, all peptides have their intrinsic properties regarding stability, and for many proteins it is impossible to define a perfect peptide. After defining peptides, these are synthesized and analysed by MS to build a spectral library. As not all peptides are stable, different parts of proteins under study might behave in different ways due to proteostasis processes, PTMs or proteoforms it is desirable to analyse multiple peptides per protein. This design should take sequence isoforms and peptide specificity for protein into account, as many proteins share redundant sequences.

There are multiple methods available, with selected reaction monitoring (SRM) and parallel reaction monitoring (PRM) being the so called “classical” methods [394]. In both precursor peptide ion specified in an inclusion list is isolated from other ion species with a quadrupole mass analyser (Figure 5.). SRM is typically done with a triple quadrupole instrument where the second quadrupole acts as a fragmentation cell, and the third scans through selected list of product ions. PRM in turn was first described for Orbitrap instrument where the precursor peptides are isolated with quadrupole and an optional MS1 scan is performed [395]. Isolated peptide precursors are fragmented in higher energy collisional dissociation or collision-induced dissociation (CID) mode, followed by simultaneous measurement of all product ions in Orbitrap. These experiments are also possible in quadrupole time of flight instruments. No matter the technique, the measured product ions are matched against the spectral library, where the ions should match library ions and their relative ion intensities, and the final quantification is done with product ions that do

not show interference from other peptides. Both PRM and SRM have their limit of quantification (LOQ) in the attomole scale and have similar quantitative accuracies [394, 396-398]. However, SRM suffers from co-isolation of interference ions due to relatively low resolution of mass selection (1-2 m/z) with each quadrupole [396], which are potentially difficult to detect in a large scale experiment. For these reasons we chose PRM for our targeted validation method (Paper II).

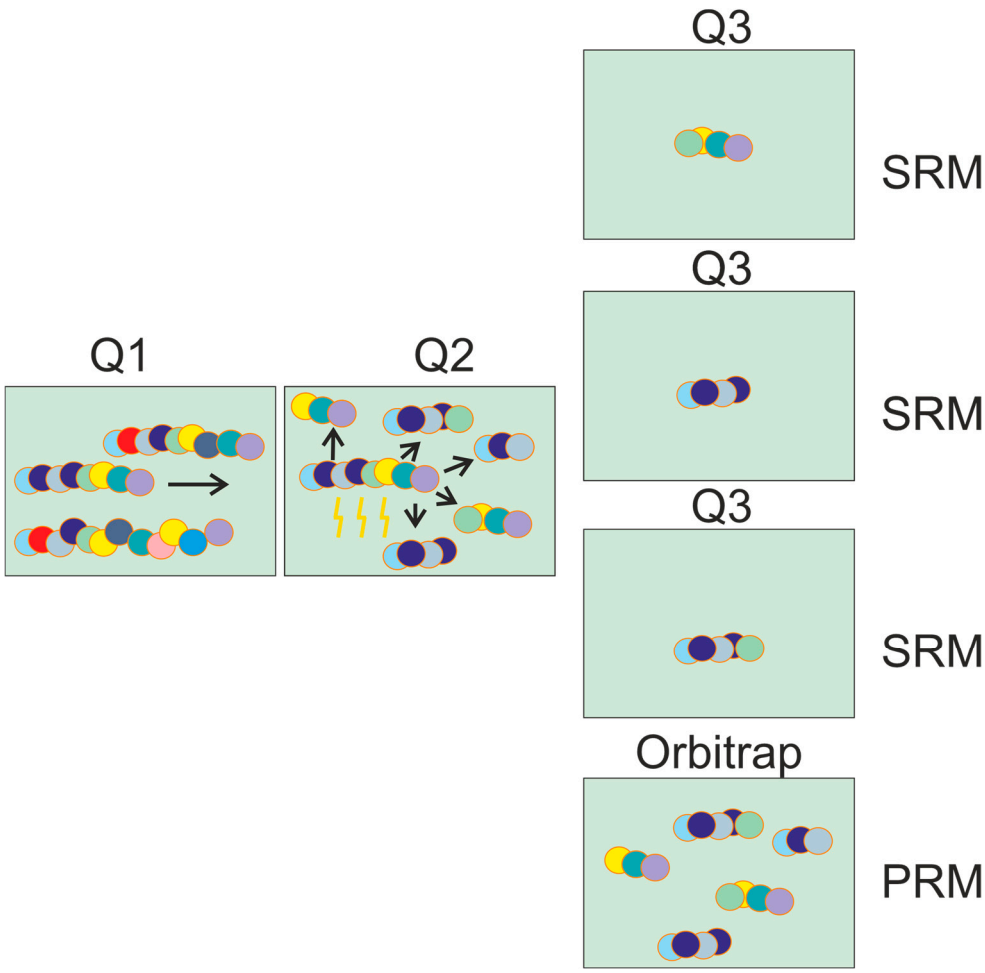


Figure 5. SRM and PRM are both targeted proteomics methods. In both methods, the isolation cell, which is typically a quadrupole are used to isolate precursor ions. In both methods second cell acts as fragmentation device(Q2), which is a second quadrupole in SRM, and fragmentation trap in Orbitrap instruments. In SRM product ions are detected one at a time in Q3, typically quadrupole. In PRM all product ions are detected in parallel in mass analyser, which is typically an Orbitrap.

In order to increase the number of peptides quantified it is possible to schedule the isolation list, so that only a subset of the isolation list is being monitored at any given time. Even with this approach, the feasible number of monitored precursor peptides is in low hundreds and requires stable chromatographic conditions and careful scheduling of the peptides. As it is unlikely that retention times replicate reproducibly over time especially with "in-house" made liquid chromatography columns, the concept of index retention time peptides was introduced by the MacCoss laboratory [399], to help adjust for retention time changes occurring over time or change of chromatographic system and gradient. An alternative approach to this is to perform a real time retention time calibration [400], which allows for tight retention time scheduling with real time MS2 matching.

As a further development to targeted methods it is possible to spike samples with a stable heavy isotope synthetic peptide (SIL) and use the mass spectrometer to monitor presence of these peptides for a triggered detection of the endogenous peptide. This was first described in index-ion triggered MS2 ion quantification (iMSTIQ) [401]. Here the sample peptides are labelled with a +4 Da mTRAQ reagent, and SIL reference peptide with light amine reactive tag. This essentially creates isobaric ion species compromised of the labelled sample and spike in peptide, which MS2 detection is triggered when a third species of peptide: heavy labelled SIL index ion peptide is detected by MS1. An advantage of this method is the fact that the method is more sensitive than inclusion list-based methods with roughly 2.5x more identified peptides at a 0.75 fmol peptide concentration [401]. The index peptide in theory should trigger the scan of the target when it is eluting, and as reference and sample peptide are simultaneously isolated, the resulting data is easy to normalize between runs. However, labelling scheme is more expensive and difficult to implement than a label free targeted method, and endogenous peptides can cause a false positive MS2 trigger from the index ion scan.

In an alternative approach, the SIL peptide is used directly to trigger the scan of endogenous peptide. Internal Standard Triggered-Parallel reaction monitoring (IS-PRM) uses SIL peptides, which are monitored with a minimum intensity threshold in scheduled mode. If the MS1 signal with a correct m/z exceeds a minimum intensity threshold, a fast low resolution scan is triggered, and resulting MS2 scan triggers a quantitative mode if it matches reference spectra [397]. Quantitative mode continues scanning the endogenous peptide with long ion accumulation times and high resolution for the duration of the scheduled time window. This enables a high specificity and sensitivity superior to both PRM and SRM, allowing more robust quantification of peptides in low scale of attomoles compared to PRM and SRM [397]. A major disadvantage of this method is that it requires programmatic access to the mass spectrometer, making the adaptation challenging in most labs. However, the SureQuant method from ThermoFisher available on newer Orbitrap instruments works on similar principle to IS-PRM and has been used to detect low abundance pTyr peptides successfully [402]. A related technique combining both iMSTIQ and

IS-PRM utilizes TMT labelling of endogenous peptides and isobaric labelling. Here the samples are mixed with a co-eluting trigger peptide tagged with alternative TMT tag, which is monitored and sequenced in real time, followed by MS2 identification, and MS3 level quantification of endogenous peptides upon reference spectra match [368]. This method and SureQuant are available only on Thermo-Fisher's tribrid series instruments, limiting its use.

Comparison of acquisition methods

Multiple studies have compared isobaric labelling, label free DDA and DIA, and the features are summarized in table 1. These studies commonly use a carrier proteome, which is supplemented with a known amount of a spike in proteome or peptides from another species with minimal proteome and sequence identity with the carrier proteome. However, these kinds of studies can introduce bias due to different tools, and unequally optimized MS conditions being used. Results from these studies consistently show that label free DDA suffers from a higher number of number of missing values and lower precision compared to isobaric labelling exemplified by the following studies [403-406], while in some cases still maintaining slightly better accuracy [403, 404]. DDA-LFQ is useful for PTM analysis where the sample complexity is not expected to be high, and number of replicates are sufficient to overcome problems associated with missing values. DDA-LFQ is also useful for open search methods, for finding unknown PTMs from samples [407]. However, until recently, this technique didn't achieve same level of proteome coverage in the same analysis time, as DIA or isobaric labelling leading to increasing popularity of DIA and isobaric labelling.

A comparison of DIA with LFQ-DDA and isobaric labelling is difficult, as DIA analysis software is still advancing rapidly, and is not as mature. Compared to LFQ-DDA, data independent acquisition shows accuracy, as well as higher proteome coverage [408, 409] and precision [408-410]. Originally DIA was producing less identifications, but higher accuracy compared to isobaric labelling [411]. This results can be partially attributed due to lower number of identifications with older DIA methods and analyses. Authors in a recent study with multiple MS search engines in complex samples found similar number of identifications for both DIA and isobaric labelling, and showed that library free DIA was marginally superior to isobaric labelling based on accuracy and precision [412]. Isobaric labelling historically achieved high proteome coverage in less time, and the results more easily scrutinized due to relative simplicity of the resulting spectra compared to DIA. However, this technique also suffers from high value of missing values between isobaric plexes. Furthermore, the combination of modern DIA software with modern instruments allows for high-throughput experiments where several thousand proteins are quantified for low tens of hundreds of samples per day analysed [413, 414]. With longer separations; >60 minute gradients, DIA is nearing full proteome coverage in terms of expressed genes [413, 415] similar to isobaric

labelling. However, the resulting DIA spectra are not identified without complicated machine learning methods, which makes human based quality control on these spectra difficult. The choice of MS analysis method for discovery depends on the application, and no one method is a perfect solution.

The advantages of targeted methods were largely discussed earlier. In the end targeted methods are used to quantify samples with a predefined and well characterized panel of targets. Ideally the limit of detection and the LOQ are defined for target peptides, as well as chromatographic behaviour. Targeted panels can be used to assess known disease specific panels in patient cohorts or validate results from discovery experiments. The separation methods for mass spectrometry are usually kept short to allow high throughput and cost is determined by whether or not the quantification is done with relative quantification or in absolute values using highly purified calibration standard peptides.

Method	Precision	Accuracy	Time	Cost
DDA-LFQ	+	++	low-medium	low
Isobaric-labelling	++	+/++/+++	Low - high	high
DIA	++	+++	low-medium	low
Targeted	+++	++++	low	low-medium

Table 1. Comparison of MS data acquisition modes. Precision and accuracy are scored with "+" where higher count indicates better performance. Time and cost are indicated categorically from low to high.

4.1.3 Tandem MS data analysis

Data analysis is a field heterogeneous field as the sample preparation and needs to be discussed in this thesis. Quantitative accuracy and precision can vary between MS analysis software platforms, which are varied and difficult to benchmark. Furthermore, data normalization and imputation can have a significant impact on quality of the output data, depending on the degree of missing values. Below I will discuss briefly how the generated spectra are assigned an identity, how data normalization is carried out and how intensity is defined and give a quick overview of the relevant statistics.

Matching MS data to identity

MS analysis generates tens of thousands of MS/MS spectra per run. Although a human is able to sequence the data by hand in simple DDA and targeted cases, the number of the spectra makes analysis impractical and specialized algorithms are used. In the simplest form of untargeted analysis, the peptide analysis software is

given a protein sequence database of the organism of interest. Next the software will perform an in silico digest of the protein sequences to give peptides and then generate a predicted spectrum, with specified PTM and digestion rules. Typically, this takes into account, the type of instrument used, and the kind of parent ion fragmentation method used such as CID. Following this, each spectrum is matched against the in silico generated spectra, and the highest scoring spectrum to database match is assigned for each spectrum [416]. In order to control for false positive matches, the analysis software will in parallel search a decoy database containing reverse, mutated, or random sequences in parallel [416]. This allows the specification of a score cut-off value, defining the number of false positives that are acceptable.

As peptide fragmentation occurs at variable efficiency at each peptide bond and the generated fragment ions do not ionize with equal efficiency, spectra have fragment ions with differing intensities. Traditional DDA search engines such as MaxQuant [111] and Mascot [417] did not take this in into account, the engines matched product ions to in silico-digested spectra irrespective of their relative intensities. As discussed earlier complex spectra such from DIA data, or targeted methods use spectral libraries from real observed peptide to spectrum matches to provide high quality and almost unambiguous peptide to spectrum matches, which takes relative peptide fragment intensities into account during spectra to peptide matching. This is particularly useful for targeted methods, where a predefined list of peptides is scanned in short LC-MS analyses. For discovery studies with DIA this can be limiting, as the best sensitivity is achieved with project specific libraries, which are time consuming to make. In addition, various sample specific features as PTMs, peptide indicative of proteoforms and sample/condition specific proteins can be left out, without very extensive peptide library building. Several tools for in silico generating spectral libraries exist such as Prosit [418], which can be utilized with modern MS search engines such as MaxQuant [388], or they are built in to search engine like in DIA-NN [389] and Spectronaut [390].

MS data pre-processing and differential expression analysis

After MS2 analysis and protein quantification is done, the data is typically filtered with missing values greater than the user set threshold, which is typically set to total number of missing values exceeding missing value % or require that at least one or more sample groups have at least certain % non-missing values. Some search engines can be set to perform protein and/or peptide level normalization, or the normalization can be performed outside the search engine. As isobaric label and targeted data normalization was discussed earlier, I will concentrate on label free discovery mode experiments here and give examples of normalization and imputation strategies. Multiple comparison studies exist comparing differential protein levels [419-421], but the choice of combination should always depend on the dataset. Of particular interest is the recent benchmark study from Lin *et al.*,

where the authors compared normalization and quantification methods from RNA-seq field to commonly used proteomics methods [421], which suggests ROTS [115] as the most suitable for differential along with statistical tests from RNA-seq abundance estimator for proteomics.

As mentioned earlier proteomic datasets typically contain many missing values, especially at the peptide level, which can be problematic for statistical testing. Missing values can be missing either completely or at random (MCAR or MAR respectively), or they can be missing not at random (MNAR). MAR refers to technical issues such as stochastic sampling in label free and isobaric labelling DDA, or incorrect peak detection. MNAR generally refers to abundance dependent missing values, due to factors like instrument detection limit or the peptide being not present in the sample or group. These are the predominant source of missing values in label-free non-DDA proteomics, although the proportion is difficult to assess and was benchmarked recently by Jin *et al*. and Lin *et al.* [421, 422]. A computational approach for determining the type of missing value present in the data is presented in MsImpute, which determines the type of missing values present in the dataset and proceeds to impute [114]. MAR values can be imputed by estimating the value from similarly behaving samples with methods such as random forest, where the observed data points are used to predict missing values [423] or k nearest neighbours, which estimates imputed values from mean of similarly behaving features [424]. As MNAR are mostly on the left side of abundance distribution, the so-called left-censored methods are used to replace missing values that are close to detection limit or below it. Examples include Perseus replacing missing values using a normal distribution [109], or simply replacing missing values with lowest sample or dataset value [425]. No universal approach exists and ideally the type of data should determine the imputation method with experimental conditions being considered.

The simplest protein normalization method is the top N method, where the top N most intense ions are averaged or median centred to provide the protein level quantification, which is biased for smaller proteins [426]. Another disadvantage of this approach is that it neglects the rest of the peptide data, which contains useful quantitative data and the underlying batch effects might be missed. Another simple method divides the sum of observed peptide intensities by number of theoretically observable peptides, and is called intensity-based absolute quantification (iBAQ) [427]. This method allows for absolute quantification by calculating protein abundance with linear regression from known external protein standard amounts. However, iBAQ is sensitive to presence of non-proteotypic peptides [428], which is problematic in bottom-up proteomics where many peptides are shared between proteins. Additionally, since the raw intensity values, which are normalized are proportional to error prone sample loading, the best accuracy iBAQ values are obtained with second round of normalization [429]. Another way of normalizing proteomics data is to first explore multiple normalization methods, and then choose

with technical replicates and known biological or technical variants, the normalization method which works best for the data, and the approach is implemented in tools like Normalyzer and the updated implementation NormalyzerDE [430, 431] and Tidyproteomics [432]. The advantage of this approach is the ability to remove researcher bias in choosing the correct method, the disadvantage is that many potentially useful normalization methods are not included depending on the tool used.

Some normalization and protein quantification methods work as a pipeline, examples being, MSstats and MSqRoB. MSstats, which is a popular proteomics tool for proteomics normalization and quantification allows the user to choose between top 3, top N and all informative peptide features for quantification [433, 434]. Peptide intensities are median, and quantile centred or normalized to specified standard proteins [433, 434]. Tukey's median polish method or linear mixed effect modelling are then used for final protein quantification [434]. Linear mixed effects are also used in MSqRoB and are used to calculate protein intensities from peptide level after choosing peptide summarization method [435, 436] like MSstats. Linear mixed effects models are complicated statistical models, which take fixed and random elements to predict a feature. Examples of fixed effects are sample condition or peptide sequence, while random effects represent subject response or run specific variance. MSqRoB requires user to determine the random and fixed effects, while MSstats 4.0 automatically determines these, making the approach more user-friendly to individuals without any special statistical knowledge [434].

MaxLFQ is one of the most popular protein level normalization methods [352, 388], which is utilized in many modern search engines such as MaxQuant [111, 388], DIA-NN [389], Spectronaut [390] and Fragpipe [391, 392, 437]. Baseline maxLFQ calculates pairwise ratio between same peptide between all samples. Next, maxLFQ calculates median of peptide ratio for each protein for each sample pair, which efficiently removes the effect of outlier peptides in robustly detected proteins. Following this maxLFQ performs calculates intensity profile with a least square analysis to make an intensity profile that fits the pairwise protein ratios the best. The original implementation was for DDA data, but maxLFQ since has been adopted for DIA in the MS search engines using maxLFQ mentioned above, and as a standalone R package [438].

In the past, many proteomic experiments were relatively small scale with less than one hundred samples. However, recent advances in peptide separation and MS instrumentation have made large-scale experiments viable economically and timewise. Ideally, a large proteomics experiment should contain balanced mix of sample conditions and a digestion standard. Batch correction methods are designed to fix technical variation coming from various sample preparation parameters such as digestion batches and time related drift. As an example, the batch correction tool proBatch first quantifies batch effects coming from the experimental design, which can include discrete effects from digestion batch, the year sample was collected or

continuous over time effects, such as performance drift with MS run batches [113]. This is followed by normalization if the data has not been normalized. If significant batch effects are encountered with normalized discrete batch effect such as digestion artefacts can be corrected with median or mean centring, or alternatively with ComBat [439], which is a Bayesian mean centring method. Recently released web application BIRCH provides the same functionality, with a more user-friendly interface [440]. Another example of batch correction method is the removal of unwanted variation (RUV-III) [118], where the variance of the least changing proteins is considered unwanted variation and are estimated technical replicates. Subsequently RUV-III minimizes the unwanted variance by subtracting the unwanted variance from stable proteins and technical replicates. Quality and overfitting of batch correction can be visualised by clustering methods such as a heatmap. Here the diffusion of discrete effects and clustering of technical repeats or standard samples is indicative of a good fit, and diffusion of standard samples shows overcorrection.

Many of the quantification and normalization methods discussed above have the aim of singling out differentially expressed proteins using correct expression quantification. However, modern high depth bottom-up MS analyses produce quantitative information from low tens to close to a hundred per protein group. This creates a problem, as peptides can come from multiple proteoforms, and summarizing peptide intensities from same gene in to one protein can mask quantitative differences that are present in the dataset [441]. There is a clear trend for protein groups with large number of identified peptides to have the least statistically significant differences, which is most likely caused by the larger number of proteoforms averaging out the differences when the peptide information is summarized to one value [441]. One approach is to use a normalization method that does robust normalization of peptide intensities, such as recently developed directLFQ [442] or to normalize the data with a tool like Normalyzer or Tidyproteomics. Performing statistical tests on the normalized data reveals peptide specific differences, but provides no information on proteoforms, or assesses if the detected peptide feature is random noise. Assigning peptides from high depth proteomic experiment to specific proteoforms requires statistical approaches such as assigning co-correlating peptides to a proteoform group as presented by Bludau and colleagues [443], or by using co-varying physical properties of the peptides to find proteoforms [444, 445].

4.2 Other methods for measuring protein level information.

4.2.1 mRNA

Steady state in omics studies refers to a cellular state where a large proportion of the protein and/or mRNA levels are stable [446] and which are then made unstable for example by stress, stimulus or differentiation leading to non-steady state. While typically under steady state highly abundant proteins have highly abundant mRNA and there is positive correlation between mRNA and protein levels, typically reports show that under steady state mRNA levels explain ~10-65% or lower of protein level variance [354-356, 384, 427, 447-453]. Discordance between mRNA and Protein level variance has been observed to be determined by protein synthesis rate from the mRNA [427, 447], by protein degradation pathways [447, 451], post-transcriptional regulation [451, 452] and sequence [452]. Although these models do not agree with each other completely they have been used in attempt to predict protein levels from mRNA.

Under the assumption that protein synthesis rate was a key component in predicting protein levels from mRNA, two independent groups observed that protein to mRNA ratio (PTR) remained constant in 12 tissues [354] and in 11 tissues and 9 cell lines [450]. With this observation, it was assumed that each protein has a transcript specific protein synthesis rate, and PTR served as a proxy constant for this. After taking median or average PTR across study tissues and cells, both groups could demonstrate that over 80% of the protein level variation could be predicted by mRNA levels. However, this approach has been questioned for its mathematical basis and for ignoring various pathways which decouple protein levels from mRNA levels [451, 454], leading to poor within gene correlations.

Additionally, in some tissues the top abundant mRNA and proteins have a low to moderate overlap depending on the cut-off in use, and sometimes the protein product for an expressed mRNA cannot detected or highly expressed protein does not have corresponding highly expressed mRNA [356, 384]. This means that some of the information leading to a phenotype is completely lost if only mRNA is considered. There have been recent improvements in predicting protein levels from mRNA based on sequence features [452], and this has been used to demonstrate that most of the explained protein level variability in the model is coming from mRNA levels [453]. Taken together, protein levels under steady state generally correlate positively with their mRNA level. When comparing different tissues and cells, or biomaterials from different individual's mRNA levels can produce different responses at protein levels, which is problematic when studying a disease phenotype, necessitating the use of proteomics to study protein level variability.

Under non-steady state conditions such as cell differentiation stress or stimulus response, the steady state dynamics between proteins and mRNA can change drastically [446]. Stress conditions can cause part of the transcriptome to be sequestered into stress-granules, where they remain inactive before they are needed or degraded. Stress can also slow or inactivate translational machinery by post-translational modifications on proteins and direct the translational machinery to synthesise proteins that are needed to survive and overcome the stress. Alternatively, a stimulus can activate translational machinery to make more protein from a steady state pool of mRNA. Increase in mRNA levels upon external stimulus or stress itself might not be immediately translated in to increased protein levels. Finally, proteins are modified post translationally, kept active/inactive by protein-protein interactions, which cannot be inferred from mRNA levels.

4.2.2 Antibody arrays and next generation methods

Although mass spectrometry remains the most mature and unambiguous method for measuring the proteome, the instrumentation is expensive and requires skilled personnel to operate. To overcome this burden several methods have been developed that try to overcome this. These methods combine techniques used in microarray, microscopy and next generation sequencing with protein level measurements. Enzyme linked immunosorbent assay and other antibody-based assays targeting only handful of targets at a time will not be discussed separately here.

Antibody arrays combine specificity of monoclonal antibodies with microchip array. The downside of antibody array is off-target binding and proteotypicity. Even if antibody has a high affinity for its target, it can still recognize other protein with similar amino-acid composition or a 3D fold, even with modern recombinant antibodies [49], and as discussed earlier the same gene can produce multiple proteoforms, which is largely missed by any antibody based platform.

As a further development to antibody arrays, a technique using modified proximity ligation was developed and commercialized as O-link [455]. In proximity ligation, two antibodies conjugated with complementary oligonucleotides are used for detecting the target protein. If both antibodies bind to same protein they come to close proximity and oligonucleotides are able to hybridize, forming a circular DNA. This can be amplified and detected with a microscope using labelled fluorescent oligonucleotide dye complementary to the antibody oligonucleotide sequence. O-link is modified so that each antibody pair has a unique tag sequence, which can be detected with quantitative real time PCR or next generation sequencing. Competing technique from Somalogic: SomaScan uses aptamers to bind target proteins [456].

Both techniques can measure thousands of protein targets, and since they are targeted, the presence of highly abundant proteins in challenging biofluids such as plasma does not interfere with the detection of more low abundant proteins, allowing for more in depth characterization. The downside of these techniques is that the affinity-based methods are an indirect readout of protein levels and off-target binding remains a risk and proteoform specific information is largely lost or averaged out. Furthermore despite having high precision, these techniques do not always correlate with each other or other assays at protein level [457-459], meaning that any result obtained needs to be scrutinized with a targeted MS method or a complementary method targeting same epitopes.

Several next generation methods are under development and measure single molecules. Nautilus Biotechnology's Nautilus platform immobilizes single protein molecules to a landing bed on an array chip with billions of binding sites. Next samples are treated sequentially through hundreds of cycles, with fluorescent probes recognizing short epitopes of protein, and using unique binding pattern and machine learning proteins are sequenced [460]. In theory, this allows for single molecule scale sequencing of a whole proteome and is not possible with any protein-based methods currently in use. As random off target binding with these kinds of reagents is expected, each epitope is sequenced multiple times, and the machine-learning algorithm calculates the probability of the binding event.

Another technique in development utilizes a modified Edman degradation and microscopy [461]. Here digested samples are immobilized on a chip, and specific amino acids are labelled with fluorescent dyes. Peptides undergo multiple Edman degradation cycles [304], and whenever a modified amino acid is removed the from the protein, the event is visible on microscope. Consequently, protein identity and quantity are established from the pattern specific to each protein and how many times each peptide has been sequenced.

Conclusions

Subjects covered in thesis cover proteomics, kinases and neurodegenerative disease Parkinson's disease. As science, moves forward in understanding complex diseases and their mechanisms traditional low throughput experiments are insufficient to explain diverse changes and networks behind the disease progress. This has heralded the age of omics, which seek to understand how biological entities work at molecular level leading to a phenotype. Proteomics is the field of omics that seeks to understand how multiple proteins and their functions lead to a certain phenotype.

The disease that is dissected in this thesis is Parkinson's disease, which is the second most common neurodegenerative disease in the world. The disease causes progressive motor symptoms and various non-motor symptoms, with no known cure. Various mechanisms including mitochondrial dysfunction, altered proteostasis and inflammation. These processes have overlapping signalling pathways and proteins involved, but it is not clear where does the pathogenesis start, and what is the main driver of the disease progression at cellular level.

One of the most commonly inherited risk factors for Parkinson's disease are in gene coding for LRRK2. LRRK2 is a large multidomain protein kinase with a second catalytic domain GTPase activity. The mutations associated with PD increase the kinase activity. In paper I we found that LRRK2 interacts with ribosomes. Upon further investigation, we demonstrate that LRRK2 kinase activity negatively regulates protein synthesis in rotenone model of PD in both vivo and vitro, and in patient primary fibroblast cells protein synthesis is specifically downregulated in LRRK2 G2019S and sporadic PD. We also show altered protein synthesis at the level of phosphoproteome in two rotenone treated rat brain regions: striatum and midbrain.

This lead to the study in paper II. Here we perform an unbiased proteomics screen to identify proteins, which synthesis is under active regulation in PD sporadic and G2019S mutation fibroblast cells. After identifying the proteins we designed a targeted peptide panel for the significantly changing proteins we identified in the proteomics screen. The resulting panel showed clear overlap between sporadic and G2019S form of PD, and proteins regulating proteostasis are downregulated. This study is has the limitation in small sample number, which requires further investigation in a larger cohort to validate the results.

Mass spectrometry based proteomics is currently the most versatile and information rich approach to proteomics. However, there is no gold standard approach within this branch of analysis and careful study design is needed to for each project. In this thesis I presented what to consider and what are the challenges in sample preparation, data acquisition and in a proteomics studies in general, and presented a brief overview of newly synthesized proteome analysis, which is utilized in paper II.

In this thesis, special emphasis was places on protein kinases, which are readily druggable master regulators of cellular signalling pathways and metabolism by modifying proteins by phosphorylation. I present a quick overview on tools to dissect complex phosphoproteomics data and discuss what are the limitations with these tools. In paper III I describe a tool to process datasets from phosphoproteomic experiments: PhosPiR [112]. This allows for deciphering of complex data and identification of kinase activity, to put their signalling in to meaningful context. In short finding the consequence of protein kinase signalling.

In conclusion, the studies presented and performed in this thesis scratch the surface of protein function in Parkinson's disease, proteomics and how kinase signalling can be studies with omics methods.

Acknowledgements

I would like to offer my thanks to my main supervisor and co-supervisor Peter James and Eleanor Coffey. Peter is a great teacher, and I'm grateful the opportunity get to know you and your stories, and many thanks for all the ideas you let me try in the lab. Things did not go as planned and life took some interesting turns, but I'm glad that you are around to see this to end. Also special thanks to Eleanor. I started working with Parkinson's disease in Eleanor's lab a long time ago, while still an undergraduate student. What had started as a summer internship in Turku, Finland evolved in to a long-term project leading to Lund, Sweden to Peter's lab. Things have not always worked optimally, but I'm glad for the trust you place in me. I'm grateful for both of you for all the patience, support and ideas along the years, and a familiar place to return, when things were tough.

I would like to thank all the wonderful co-worker's and collaborators who have supported me during the years. From Lund Fredrik Levander, Emila Kurbasic, Elena Goltseva, Valentina Siino, Leena Liljedahl, Mats Mågård and Karin Hansson. I would like to thank Fredrik, Elena, Valentina, Leena and Mats for all the help and chats. Special thanks to Emila and Karin. Emila for nice chats in and out of the office. Karin for patience and support with the aging MS instrumentation at the department. I had very challenging samples, that were close to impossible to clean up, and causing problems with the instrument. You never got angry and supported me in my projects.

Also I would like to thank everyone who I've worked with in Turku. I wish special thanks to Prasannakumar Deshpande, Patrik Hollos, Ye Hong, Ali Mostafa, Arttu Heino, Veronica Fagerholm and Jelena Gnjatovic. I've known Prasanna as long as I've worked in science. First as a mentor and later as a friend. I'm glad that you are working again with me even if I've been temproraily occupied with non-work related issues. Patrik Hollos for all the adventures we had outside the work and in foreign trips. Ye Hong for friendly discussions and collaborations, I may sometimes be thick, but you've always listened and explained things carefully to me. Ali Mostafa has not been around for long, but has already given many ideas and help. Arttu and rest of the proteomics core have helped me to go through challenging times and taught me things about MS instrumentation in areas where I have gaps in knowledge. Veronica for help with the most challenging +1200 clinical cell sample project I've done so far. If it was not for you I probably would have been crushed

by the lab work before I even got to the instrument. Finally thanks for Jelena for helping to finish up the Parkinson's study that started many years ago.

Finally, I would like to thank my wife Tia and daughter Irina. Things have been difficult for us, but we have survived despite the obstacles we've encountered. The future still looks bright, good things will come and new adventures will follow.

27th of March, 2024

Dani Flinkman

References

1. Stovner, L.J., et al., *Global, regional, and national burden of migraine and tension-type headache, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016.* The Lancet Neurology, 2018. **17**(11): p. 954-976.
2. Wallin, M.T., et al., *Global, regional, and national burden of multiple sclerosis 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016.* The Lancet Neurology, 2019. **18**(3): p. 269-285.
3. Olesen, J., et al., *The economic cost of brain disorders in Europe.* Eur J Neurol, 2012. **19**(1): p. 155-62.
4. Zolnierowicz, S. and M. Bollen, *Protein phosphorylation and protein phosphatases. De Panne, Belgium, September 19-24, 1999.* EMBO J, 2000. **19**(4): p. 483-8.
5. Olsen, J.V., et al., *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.* Cell, 2006. **127**(3): p. 635-48.
6. Sharma, K., et al., *Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling.* Cell Rep, 2014. **8**(5): p. 1583-94.
7. Hunter, T. and B.M. Sefton, *Transforming gene product of Rous sarcoma virus phosphorylates tyrosine.* Proc Natl Acad Sci U S A, 1980. **77**(3): p. 1311-5.
8. Chen, C.C., et al., *Occurrence and distribution of acid-labile histone phosphates in regenerating rat liver.* Biochemistry, 1974. **13**(18): p. 3785-9.
9. Wagner, P.D. and N.D. Vu, *Phosphorylation of ATP-citrate lyase by nucleoside diphosphate kinase.* J Biol Chem, 1995. **270**(37): p. 21758-64.
10. Wakim, B.T. and G.D. Aswad, *Ca(2+)-calmodulin-dependent phosphorylation of arginine in histone 3 by a nuclear kinase from mouse leukemia cells.* J Biol Chem, 1994. **269**(4): p. 2722-7.
11. Wagner, P.D. and N.D. Vu, *Histidine to aspartate phosphotransferase activity of nm23 proteins: phosphorylation of aldolase C on Asp-319.* Biochem J, 2000. **346 Pt 3**(Pt 3): p. 623-30.
12. Cieśla, J., T. Frączyk, and W. Rode, *Phosphorylation of basic amino acid residues in proteins: important but easily missed.* Acta Biochim Pol, 2011. **58**(2): p. 137-48.
13. Attwood, P.V., P.G. Besant, and M.J. Piggott, *Focus on phosphoaspartate and phosphoglutamate.* Amino Acids, 2011. **40**(4): p. 1035-51.
14. Hardman, G., et al., *Strong anion exchange-mediated phosphoproteomics reveals extensive human non-canonical phosphorylation.* EMBO J, 2019. **38**(21): p. e100847.

15. Leijten, N.M., A.J.R. Heck, and S. Lemeer, *Histidine phosphorylation in human cells; a needle or phantom in the haystack?* Nat Methods, 2022. **19**(7): p. 827-828.
16. Knighton, D.R., et al., *Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase.* Science, 1991. **253**(5018): p. 407-14.
17. Taylor, S.S. and A.P. Kornev, *Protein kinases: evolution of dynamic regulatory proteins.* Trends Biochem Sci, 2011. **36**(2): p. 65-77.
18. Nolen, B., S. Taylor, and G. Ghosh, *Regulation of protein kinases; controlling activity through activation segment conformation.* Mol Cell, 2004. **15**(5): p. 661-75.
19. Iakoucheva, L.M., et al., *The importance of intrinsic disorder for protein phosphorylation.* Nucleic Acids Res, 2004. **32**(3): p. 1037-49.
20. Heo, Y.S., et al., *Structural basis for the selective inhibition of JNK1 by the scaffolding protein JIP1 and SP600125.* EMBO J, 2004. **23**(11): p. 2185-95.
21. Schubert, A.F., et al., *Structure of PINK1 in complex with its substrate ubiquitin.* Nature, 2017. **552**(7683): p. 51-56.
22. Kornev, A.P., et al., *Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism.* Proc Natl Acad Sci U S A, 2006. **103**(47): p. 17783-8.
23. Kornev, A.P., S.S. Taylor, and L.F. Ten Eyck, *A helix scaffold for the assembly of active protein kinases.* Proc Natl Acad Sci U S A, 2008. **105**(38): p. 14377-82.
24. Huse, M. and J. Kuriyan, *The conformational plasticity of protein kinases.* Cell, 2002. **109**(3): p. 275-82.
25. Creixell, P., et al., *Unmasking determinants of specificity in the human kinome.* Cell, 2015. **163**(1): p. 187-201.
26. Manning, G., et al., *The Protein Kinase Complement of the Human Genome.* Science, 2002. **298**(5600): p. 1912-1934.
27. Eid, S., et al., *KinMap: a web-based tool for interactive navigation through human kinome data.* 2017. **18**(1): p. 16.
28. Berginski, M.E., et al., *The Dark Kinase Knowledgebase: an online compendium of knowledge and experimental results of understudied kinases.* Nucleic Acids Res, 2021. **49**(D1): p. D529-D535.
29. Moret, N., et al., *A resource for exploring the understudied human kinome for research and therapeutic opportunities.* bioRxiv, 2021: p. 2020.04.02.022277.
30. Yang, W., et al., *PKM2 phosphorylates histone H3 and promotes gene transcription and tumorigenesis.* Cell, 2012. **150**(4): p. 685-96.
31. Consortium, U., *UniProt: the universal protein knowledgebase in 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D480-D489.
32. Tunyasuvunakool, K., et al., *Highly accurate protein structure prediction for the human proteome.* Nature, 2021. **596**(7873): p. 590-596.
33. Tomoni, A., et al. *Pseudokinases: From Allosteric Regulation of Catalytic Domains and the Formation of Macromolecular Assemblies to Emerging Drug Targets*. Catalysts, 2019. **9**, DOI: 10.3390/catal9090778.

34. Mukherjee, K., et al., *CASK Functions as a Mg2+-independent neurexin kinase.* Cell, 2008. **133**(2): p. 328-39.
35. Eswaran, J., et al., *Structure and functional characterization of the atypical human kinase haspin.* Proc Natl Acad Sci U S A, 2009. **106**(48): p. 20198-203.
36. Villa, F., et al., *Crystal structure of the catalytic domain of Haspin, an atypical kinase implicated in chromatin organization.* Proc Natl Acad Sci U S A, 2009. **106**(48): p. 20204-9.
37. Schmidt, S.H., et al., *The dynamic switch mechanism that leads to activation of LRRK2 is embedded in the DFGψ motif in the kinase domain.* Proc Natl Acad Sci U S A, 2019. **116**(30): p. 14979-14988.
38. Lavoie, H., J. Gagnon, and M. Therrien, *ERK signalling: a master regulator of cell behaviour, life and fate.* Nat Rev Mol Cell Biol, 2020. **21**(10): p. 607-632.
39. Ferrell, J.E., *Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs.* Trends Biochem Sci, 1996. **21**(12): p. 460-6.
40. Tedeschi, A. and F. Bradke, *The DLK signalling pathway--a double-edged sword in neural development and regeneration.* EMBO Rep, 2013. **14**(7): p. 605-14.
41. Bogoyevitch, M.A. and B. Kobe, *Uses for JNK: the many and varied substrates of the c-Jun N-terminal kinases.* Microbiol Mol Biol Rev, 2006. **70**(4): p. 1061-95.
42. Tararuk, T., et al., *JNK1 phosphorylation of SCG10 determines microtubule dynamics and axodendritic length.* J Cell Biol, 2006. **173**(2): p. 265-77.
43. Coffey, E.T., et al., *c-Jun N-terminal protein kinase (JNK) 2/3 is specifically activated by stress, mediating c-Jun activation, in the presence of constitutive JNK1 activity in cerebellar neurons.* J Neurosci, 2002. **22**(11): p. 4335-45.
44. Björkblom, B., et al., *Constitutively active cytoplasmic c-Jun N-terminal kinase 1 is a dominant regulator of dendritic architecture: role of microtubule-associated protein 2 as an effector.* J Neurosci, 2005. **25**(27): p. 6350-61.
45. Niu, J., et al., *Palmitoylation couples the kinases DLK and JNK3 to facilitate prodegenerative axon-to-soma signaling.* Sci Signal, 2022. **15**(727): p. eabh2674.
46. Sato, K., et al., *Inverse correlation between Thr-669 and constitutive tyrosine phosphorylation in the asymmetric epidermal growth factor receptor dimer conformation.* Cancer Sci, 2013. **104**(10): p. 1315-22.
47. Towbin, H., T. Staehelin, and J. Gordon, *Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications.* Proc Natl Acad Sci U S A, 1979. **76**(9): p. 4350-4.
48. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.
49. Marcon, E., et al., *Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation.* Nat Methods, 2015. **12**(8): p. 725-31.
50. Bradbury, A. and A. Plückthun, *Reproducibility: Standardize antibodies used in research.* Nature, 2015. **518**(7537): p. 27-9.
51. Sutherland, C., J.A. MacDonald, and M.P. Walsh, *Analysis of phosphorylation of the myosin-targeting subunit of myosin light chain phosphatase by Phos-tag SDS-PAGE.* Am J Physiol Cell Physiol, 2016. **310**(8): p. C681-91.

52. Grahame, H., *Protein Phosphorylation: A Practical Approach* 1999, Oxford University Press. p. 431.
53. Gafken, P.R. and P.D. Lampe, *Methodologies for characterizing phosphoproteins by mass spectrometry.* Cell Commun Adhes, 2006. **13**(5-6): p. 249-62.
54. Kinoshita, E., et al., *Phosphate-binding tag, a new tool to visualize phosphorylated proteins.* Mol Cell Proteomics, 2006. **5**(4): p. 749-57.
55. Thorsness, P.E. and D.E. Koshland, *Inactivation of isocitrate dehydrogenase by phosphorylation is mediated by the negative charge of the phosphate.* J Biol Chem, 1987. **262**(22): p. 10422-5.
56. Chen, Z. and P.A. Cole, *Synthetic approaches to protein phosphorylation.* Curr Opin Chem Biol, 2015. **28**: p. 115-22.
57. Westerlund, N., et al., *Phosphorylation of SCG10/stathmin-2 determines multipolar stage exit and neuronal migration rate.* Nat Neurosci, 2011. **14**(3): p. 305-13.
58. Roskoski, R., *Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes.* Pharmacol Res, 2016. **103**: p. 26-48.
59. Abdeldayem, A., et al., *Advances in covalent kinase inhibitors.* Chem Soc Rev, 2020. **49**(9): p. 2617-2687.
60. Beausoleil, S.A., et al., *Large-scale characterization of HeLa cell nuclear phosphoproteins.* Proc Natl Acad Sci U S A, 2004. **101**(33): p. 12130-5.
61. Zhang, K., *From purification of large amounts of phospho-compounds (nucleotides) to enrichment of phospho-peptides using anion-exchanging resin.* Anal Biochem, 2006. **357**(2): p. 225-31.
62. Nühse, T.S., et al., *Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry.* Mol Cell Proteomics, 2003. **2**(11): p. 1234-43.
63. Pinkse, M.W., et al., *Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns.* Anal Chem, 2004. **76**(14): p. 3935-43.
64. Gruhler, A., et al., *Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway.* Mol Cell Proteomics, 2005. **4**(3): p. 310-27.
65. Low, T.Y., et al., *WIDENING THE BOTTLENECK OF PHOSPHOPROTEOMICS: EVOLVING STRATEGIES FOR PHOSPHOPEPTIDE ENRICHMENT.* Mass Spectrom Rev, 2021. **40**(4): p. 309-333.
66. Zhao, M.X., et al., *Protein phosphorylation database and prediction tools.* Brief Bioinform, 2023. **24**(2).
67. Needham, E.J., et al., *Illuminating the dark phosphoproteome.* Sci Signal, 2019. **12**(565).
68. Dolgin, E., *The most popular genes in the human genome.* Nature, 2017. **551**(7681): p. 427-431.
69. Wood, V., et al., *Hidden in plain sight: what remains to be discovered in the eukaryotic proteome?* Open Biol, 2019. **9**(2): p. 180241.

70. Hornbeck, P.V., et al., *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations.* Nucleic Acids Res, 2015. **43**(Database issue): p. D512-20.
71. Deutsch, E.W., *The PeptideAtlas Project.* Methods Mol Biol, 2010. **604**: p. 285-96.
72. Sadowski, I., et al., *The PhosphoGRID Saccharomyces cerevisiae protein phosphorylation site database: version 2.0 update.* Database (Oxford), 2013. **2013**: p. bat026.
73. Kalyuzhnyy, A., et al., *Profiling the Human Phosphoproteome to Estimate the True Extent of Protein Phosphorylation.* J Proteome Res, 2022. **21**(6): p. 1510-1524.
74. Perfetto, L., et al., *SIGNOR: a database of causal relationships between biological entities.* Nucleic Acids Res, 2016. **44**(D1): p. D548-54.
75. Lo Surdo, P., et al., *SIGNOR 3.0, the SIGnaling network open resource 3.0: 2022 update.* Nucleic Acids Res, 2023. **51**(D1): p. D631-D637.
76. Krug, K., et al., *A Curated Resource for Phosphosite-specific Signature Analysis.* Mol Cell Proteomics, 2019. **18**(3): p. 576-593.
77. Peri, S., et al., *Development of human protein reference database as an initial platform for approaching systems biology in humans.* Genome Res, 2003. **13**(10): p. 2363-71.
78. Oughtred, R., et al., *The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions.* Protein Sci, 2021. **30**(1): p. 187-200.
79. Savage, S.R. and B. Zhang, *Using phosphoproteomics data to understand cellular signaling: a comprehensive guide to bioinformatics resources.* Clin Proteomics, 2020. **17**: p. 27.
80. Linding, R., et al., *Systematic discovery of in vivo phosphorylation networks.* Cell, 2007. **129**(7): p. 1415-26.
81. Parca, L., et al., *Kinome-wide identification of phosphorylation networks in eukaryotic proteomes.* Bioinformatics, 2019. **35**(3): p. 372-379.
82. Pinna, L.A. and M. Ruzzene, *How do protein kinases recognize their substrates?* Biochim Biophys Acta, 1996. **1314**(3): p. 191-225.
83. Johnson, J.L., et al., *An atlas of substrate specificities for the human serine/threonine kinome.* Nature, 2023. **613**(7945): p. 759-766.
84. Rust, H.L. and P.R. Thompson, *Kinase consensus sequences: a breeding ground for crosstalk.* ACS Chem Biol, 2011. **6**(9): p. 881-92.
85. Piersma, S.R., et al., *Inferring kinase activity from phosphoproteomic data: Tool comparison and recent applications.* Mass Spectrom Rev, 2022: p. e21808.
86. Lachmann, A. and A. Ma'ayan, *KEA: kinase enrichment analysis.* Bioinformatics, 2009. **25**(5): p. 684-6.
87. Kuleshov, M.V., et al., *KEA3: improved kinase enrichment analysis via data integration.* Nucleic Acids Res, 2021. **49**(W1): p. W304-W316.
88. Casado, P., et al., *Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells.* Sci Signal, 2013. **6**(268): p. rs6.

89. Wiredja, D.D., M. Koyutürk, and M.R. Chance, *The KSEA App: a web-based tool for kinase activity inference from quantitative phosphoproteomics.* Bioinformatics, 2017. **33**(21): p. 3489-3491.
90. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nat Genet, 2003. **34**(3): p. 267-73.
91. Mari, T., et al., *In Vitro Kinase-to-Phosphosite Database (iKiP-DB) Predicts Kinase Activity in Phosphoproteomic Datasets.* J Proteome Res, 2022. **21**(6): p. 1575-1587.
92. Sugiyama, N., H. Imamura, and Y. Ishihama, *Large-scale Discovery of Substrates of the Human Kinome.* Sci Rep, 2019. **9**(1): p. 10503.
93. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
94. Aleksander, S.A., et al., *The Gene Ontology knowledgebase in 2023.* Genetics, 2023. **224**(1).
95. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.
96. Pico, A.R., et al., *WikiPathways: pathway editing for the people.* PLoS Biol, 2008. **6**(7): p. e184.
97. Martens, M., et al., *WikiPathways: connecting communities.* Nucleic Acids Res, 2021. **49**(D1): p. D613-D621.
98. Sidiropoulos, K., et al., *Reactome enhanced pathway visualization.* Bioinformatics, 2017. **33**(21): p. 3461-3467.
99. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.
100. Szklarczyk, D., et al., *The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest.* Nucleic Acids Res, 2023. **51**(D1): p. D638-D646.
101. Leeming, M.G., et al., *Phosphomatics: interactive interrogation of substrate-kinase networks in global phosphoproteomics datasets.* Bioinformatics, 2021. **37**(11): p. 1635-1636.
102. Barshir, R., et al., *The TissueNet database of human tissue protein-protein interactions.* Nucleic Acids Res, 2013. **41**(Database issue): p. D841-4.
103. Ziv, M., et al., *The TissueNet v.3 Database: Protein-protein Interactions in Adult and Embryonic Human Tissue contexts.* J Mol Biol, 2022. **434**(11): p. 167532.
104. Greene, C.S., et al., *Understanding multicellular function and disease with human tissue-specific networks.* Nat Genet, 2015. **47**(6): p. 569-76.
105. Raaijmakers, L.M., et al., *PhosphoPath: Visualization of Phosphosite-centric Dynamics in Temporal Molecular Networks.* J Proteome Res, 2015. **14**(10): p. 4332-41.
106. Martinez-Val, A., et al., *Spatial-proteomics reveals phospho-signaling dynamics at subcellular resolution.* Nat Commun, 2021. **12**(1): p. 7113.
107. Liu, Y., et al., *Spatiotemporally resolved subcellular phosphoproteomics.* Proc Natl Acad Sci U S A, 2021. **118**(25).

108. Ressa, A., et al., *PaDuA: A Python Library for High-Throughput (Phospho)proteomics Data Analysis.* J Proteome Res, 2019. **18**(2): p. 576-584.
109. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data.* Nat Methods, 2016. **13**(9): p. 731-40.
110. Webb-Robertson, B.J., et al., *Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics.* J Proteome Res, 2015. **14**(5): p. 1993-2001.
111. 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.
112. Hong, Y., et al., *PhosPiR: an automated phosphoproteomic pipeline in R.* Brief Bioinform, 2022. **23**(1).
113. Čuklina, J., et al., *Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial.* Mol Syst Biol, 2021. **17**(8): p. e10240.
114. Hediyeh-Zadeh, S., A.I. Webb, and M.J. Davis, *MsImpute: Estimation of missing peptide intensity data in label-free quantitative mass spectrometry.* Mol Cell Proteomics, 2023: p. 100558.
115. Suomi, T., et al., *ROTS: An R package for reproducibility-optimized statistical testing.* PLOS Computational Biology, 2017. **13**(5): p. e1005562.
116. Del Carratore, F., et al., *RankProd 2.0: a refactored bioconductor package for detecting differentially expressed features in molecular profiling datasets.* Bioinformatics, 2017. **33**(17): p. 2774-2775.
117. Waardenberg and Ashley, *KinSwingR: KinSwingR: network-based kinase activity prediction*. 2023: R package version 1.18.0.
118. Molania, R., et al., *A new normalization for Nanostring nCounter gene expression data.* Nucleic Acids Res, 2019. **47**(12): p. 6073-6083.
119. Collaborators, G.N., *Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016.* Lancet Neurol, 2019. **18**(5): p. 459-480.
120. Ou, Z., et al., *Global Trends in the Incidence, Prevalence, and Years Lived With Disability of Parkinson's Disease in 204 Countries/Territories From 1990 to 2019.* Front Public Health, 2021. **9**: p. 776847.
121. Hirsch, L., et al., *The Incidence of Parkinson's Disease: A Systematic Review and Meta-Analysis.* Neuroepidemiology, 2016. **46**(4): p. 292-300.
122. Kordower, J.H., et al., *Disease duration and the integrity of the nigrostriatal system in Parkinson's disease.* Brain, 2013. **136**(Pt 8): p. 2419-31.
123. Furukawa, K., et al., *Motor Progression and Nigrostriatal Neurodegeneration in Parkinson Disease.* Ann Neurol, 2022. **92**(1): p. 110-121.
124. Kalia, L.V. and A.E. Lang, *Parkinson's disease.* Lancet, 2015. **386**(9996): p. 896-912.
125. Rodriguez-Sanchez, F., et al., *Identifying Parkinson's disease subtypes with motor and non-motor symptoms via model-based multi-partition clustering.* Sci Rep, 2021. **11**(1): p. 23645.
126. Sandor, C., et al., *Universal clinical Parkinson's disease axes identify a major influence of neuroinflammation.* Genome Med, 2022. **14**(1): p. 129.

127. Kaiser, S., et al., *A proteogenomic view of Parkinson's disease causality and heterogeneity.* NPJ Parkinsons Dis, 2023. **9**(1): p. 24.
128. Simuni, T., et al., *A biological definition of neuronal α-synuclein disease: towards an integrated staging system for research.* Lancet Neurol, 2024. **23**(2): p. 178-190.
129. Noyce, A.J., A.J. Lees, and A.E. Schrag, *The prediagnostic phase of Parkinson's disease.* J Neurol Neurosurg Psychiatry, 2016. **87**(8): p. 871-8.
130. Singleton, A.B., et al., *alpha-Synuclein locus triplication causes Parkinson's disease.* Science, 2003. **302**(5646): p. 841.
131. Siddiqui, I.J., N. Pervaiz, and A.A. Abbasi, *The Parkinson Disease gene SNCA: Evolutionary and structural insights with pathological implication.* Sci Rep, 2016. **6**: p. 24475.
132. Bandres-Ciga, S., et al., *Genetics of Parkinson's disease: An introspection of its journey towards precision medicine.* Neurobiol Dis, 2020. **137**: p. 104782.
133. Funayama, M., et al., *A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1.* Ann Neurol, 2002. **51**(3): p. 296-301.
134. Zimprich, A., et al., *Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology.* Neuron, 2004. **44**(4): p. 601-7.
135. Paisán-Ruíz, C., et al., *Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease.* Neuron, 2004. **44**(4): p. 595-600.
136. Kachergus, J., et al., *Identification of a novel LRRK2 mutation linked to autosomal dominant parkinsonism: evidence of a common founder across European populations.* Am J Hum Genet, 2005. **76**(4): p. 672-80.
137. Di Fonzo, A., et al., *A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson's disease.* Lancet, 2005. **365**(9457): p. 412-5.
138. Gilks, W.P., et al., *A common LRRK2 mutation in idiopathic Parkinson's disease.* Lancet, 2005. **365**(9457): p. 415-6.
139. Hui, K.Y., et al., *Functional variants in the LRRK2 Confer Pleiotropic Effects on Risk for Crohn's Disease and Parkinson's Disease.* Sci Transl Med, 2018. **10**(423).
140. Rocha, E.M., et al., *LRRK2 and idiopathic Parkinson's disease.* Trends Neurosci, 2022. **45**(3): p. 224-236.
141. Bardien, S., et al., *Genetic characteristics of leucine-rich repeat kinase 2 (LRRK2) associated Parkinson's disease.* Parkinsonism Relat Disord, 2011. **17**(7): p. 501-8.
142. Saunders-Pullman, R., et al., *Progression in the LRRK2-Asssociated Parkinson Disease Population.* JAMA Neurol, 2018. **75**(3): p. 312-319.
143. Fraser, K.B., et al., *Ser(P)-1292 LRRK2 in urinary exosomes is elevated in idiopathic Parkinson's disease.* Mov Disord, 2016. **31**(10): p. 1543-1550.
144. Di Maio, R., et al., *LRRK2 activation in idiopathic Parkinson's disease.* Sci Transl Med, 2018. **10**(451).
145. Deshpande, P., et al., *Protein synthesis is suppressed in sporadic and familial Parkinson's disease by LRRK2.* FASEB J, 2020. **34**(11): p. 14217-14233.
146. Gardet, A., et al., *LRRK2 is involved in the IFN-gamma response and host response to pathogens.* J Immunol, 2010. **185**(9): p. 5577-85.

147. Skelton, P.D., V. Tokars, and L. Parisiadou, *LRRK2 at Striatal Synapses: Cell-Type Specificity and Mechanistic Insights.* Cells, 2022. **11**(1).
148. Greggio, E., et al., *The Parkinson disease-associated leucine-rich repeat kinase 2 (LRRK2) is a dimer that undergoes intramolecular autophosphorylation.* J Biol Chem, 2008. **283**(24): p. 16906-14.
149. Taymans, J.M., et al., *Perspective on the current state of the LRRK2 field.* NPJ Parkinsons Dis, 2023. **9**(1): p. 104.
150. Sen, S., P.J. Webber, and A.B. West, *Dependence of leucine-rich repeat kinase 2 (LRRK2) kinase activity on dimerization.* J Biol Chem, 2009. **284**(52): p. 36346-36356.
151. Berger, Z., K.A. Smith, and M.J. Lavoie, *Membrane localization of LRRK2 is associated with increased formation of the highly active LRRK2 dimer and changes in its phosphorylation.* Biochemistry, 2010. **49**(26): p. 5511-23.
152. Leandrou, E., et al., *Kinase activity of mutant LRRK2 manifests differently in hetero-dimeric vs. homo-dimeric complexes.* Biochem J, 2019. **476**(3): p. 559-579.
153. Kett, L.R., et al., *LRRK2 Parkinson disease mutations enhance its microtubule association.* Hum Mol Genet, 2012. **21**(4): p. 890-9.
154. Blanca Ramírez, M., et al., *GTP binding regulates cellular localization of Parkinson's disease-associated LRRK2.* Hum Mol Genet, 2017. **26**(14): p. 2747-2767.
155. Myasnikov, A., et al., *Structural analysis of the full-length human LRRK2.* Cell, 2021. **184**(13): p. 3519-3527.e10.
156. Schmidt, S.H., et al., *Conformation and dynamics of the kinase domain drive subcellular location and activation of LRRK2.* Proc Natl Acad Sci U S A, 2021. **118**(23).
157. Ray, S., et al., *The Parkinson disease-linked LRRK2 protein mutation I2020T stabilizes an active state conformation leading to increased kinase activity.* J Biol Chem, 2014. **289**(19): p. 13042-53.
158. Ito, G., et al., *GTP binding is essential to the protein kinase activity of LRRK2, a causative gene product for familial Parkinson's disease.* Biochemistry, 2007. **46**(5): p. 1380-8.
159. Luzón-Toro, B., et al., *Mechanistic insight into the dominant mode of the Parkinson's disease-associated G2019S LRRK2 mutation.* Hum Mol Genet, 2007. **16**(17): p. 2031-9.
160. Li, X., et al., *Reevaluation of phosphorylation sites in the Parkinson disease-associated leucine-rich repeat kinase 2.* J Biol Chem, 2010. **285**(38): p. 29569-76.
161. Dzamko, N., et al., *Inhibition of LRRK2 kinase activity leads to dephosphorylation of Ser(910)/Ser(935), disruption of 14-3-3 binding and altered cytoplasmic localization.* Biochem J, 2010. **430**(3): p. 405-13.
162. Sheng, Z., et al., *Ser1292 autophosphorylation is an indicator of LRRK2 kinase activity and contributes to the cellular effects of PD mutations.* Sci Transl Med, 2012. **4**(164): p. 164ra161.

163. Reynolds, A., et al., *LRRK2 kinase activity and biology are not uniformly predicted by its autophosphorylation and cellular phosphorylation site status.* Front Mol Neurosci, 2014. **7**: p. 54.
164. Henderson, J.L., et al., *Discovery and preclinical profiling of 3-[4-(morpholin-4-yl)-7H-pyrrolo[2,3-d]pyrimidin-5-yl]benzonitrile (PF-06447475), a highly potent, selective, brain penetrant, and in vivo active LRRK2 kinase inhibitor.* J Med Chem, 2015. **58**(1): p. 419-32.
165. Ito, G., et al., *Phos-tag analysis of Rab10 phosphorylation by LRRK2: a powerful assay for assessing kinase function and inhibitors.* Biochem J, 2016. **473**(17): p. 2671-85.
166. Kluss, J.H., et al., *Detection of endogenous S1292 LRRK2 autophosphorylation in mouse tissue as a readout for kinase activity.* NPJ Parkinsons Dis, 2018. **4**: p. 13.
167. Fraser, K.B., et al., *Urinary LRRK2 phosphorylation predicts parkinsonian phenotypes in G2019S LRRK2 carriers.* Neurology, 2016. **86**(11): p. 994-9.
168. Nichols, R.J., et al., *14-3-3 binding to LRRK2 is disrupted by multiple Parkinson's disease-associated mutations and regulates cytoplasmic localization.* Biochem J, 2010. **430**(3): p. 393-404.
169. Deng, X., et al., *Characterization of a selective inhibitor of the Parkinson's disease kinase LRRK2.* Nat Chem Biol, 2011. **7**(4): p. 203-5.
170. Choi, H.G., et al., *Brain Penetrant LRRK2 Inhibitor.* ACS Med Chem Lett, 2012. **3**(8): p. 658-662.
171. Doggett, E.A., et al., *Phosphorylation of LRRK2 serines 955 and 973 is disrupted by Parkinson's disease mutations and LRRK2 pharmacological inhibition.* J Neurochem, 2012. **120**(1): p. 37-45.
172. Reith, A.D., et al., *GSK2578215A; a potent and highly selective 2-arylmethyloxy-5-substitutent-N-arylbenzamide LRRK2 kinase inhibitor.* Bioorg Med Chem Lett, 2012. **22**(17): p. 5625-9.
173. Fell, M.J., et al., *MLi-2, a Potent, Selective, and Centrally Active Compound for Exploring the Therapeutic Potential and Safety of LRRK2 Kinase Inhibition.* J Pharmacol Exp Ther, 2015. **355**(3): p. 397-409.
174. Perera, G., et al., *Inhibitor treatment of peripheral mononuclear cells from Parkinson's disease patients further validates LRRK2 dephosphorylation as a pharmacodynamic biomarker.* Sci Rep, 2016. **6**: p. 31391.
175. Steger, M., et al., *Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases.* Elife, 2016. **5**.
176. Fan, Y., et al., *Interrogating Parkinson's disease LRRK2 kinase pathway activity by assessing Rab10 phosphorylation in human neutrophils.* Biochem J, 2018. **475**(1): p. 23-44.
177. Dzamko, N., et al., *LRRK2 levels and phosphorylation in Parkinson's disease brain and cases with restricted Lewy bodies.* Mov Disord, 2017. **32**(3): p. 423-432.
178. Dzamko, N., et al., *Measurement of LRRK2 and Ser910/935 phosphorylated LRRK2 in peripheral blood mononuclear cells from idiopathic Parkinson's disease patients.* J Parkinsons Dis, 2013. **3**(2): p. 145-52.

179. Padmanabhan, S., et al., *An Assessment of LRRK2 Serine 935 Phosphorylation in Human Peripheral Blood Mononuclear Cells in Idiopathic Parkinson's Disease and G2019S LRRK2 Cohorts.* J Parkinsons Dis, 2020. **10**(2): p. 623-629.
180. Jaleel, M., et al., *LRRK2 phosphorylates moesin at threonine-558: characterization of how Parkinson's disease mutants affect kinase activity.* Biochem J, 2007. **405**(2): p. 307-17.
181. Lis, P., et al., *Development of phospho-specific Rab protein antibodies to monitor.* Biochem J, 2018. **475**(1): p. 1-22.
182. Fan, Y., et al., *R1441G but not G2019S mutation enhances LRRK2 mediated Rab10 phosphorylation in human peripheral blood neutrophils.* Acta Neuropathol, 2021. **142**(3): p. 475-494.
183. Thirstrup, K., et al., *Selective LRRK2 kinase inhibition reduces phosphorylation of endogenous Rab10 and Rab12 in human peripheral mononuclear blood cells.* Sci Rep, 2017. **7**(1): p. 10300.
184. Atashrazm, F., et al., *LRRK2-mediated Rab10 phosphorylation in immune cells from Parkinson's disease patients.* Mov Disord, 2019. **34**(3): p. 406-415.
185. Wang, S., et al., *Elevated Urinary Rab10 Phosphorylation in Idiopathic Parkinson Disease.* Mov Disord, 2022. **37**(7): p. 1454-1464.
186. Petropoulou-Vathi, L., et al., *Distinct profiles of LRRK2 activation and Rab GTPase phosphorylation in clinical samples from different PD cohorts.* NPJ Parkinsons Dis, 2022. **8**(1): p. 73.
187. Li, X., et al., *Leucine-rich repeat kinase 2 (LRRK2)/PARK8 possesses GTPase activity that is altered in familial Parkinson's disease R1441C/G mutants.* J Neurochem, 2007. **103**(1): p. 238-47.
188. Lewis, P.A., et al., *The R1441C mutation of LRRK2 disrupts GTP hydrolysis.* Biochem Biophys Res Commun, 2007. **357**(3): p. 668-71.
189. Guo, L., et al., *The Parkinson's disease-associated protein, leucine-rich repeat kinase 2 (LRRK2), is an authentic GTPase that stimulates kinase activity.* Exp Cell Res, 2007. **313**(16): p. 3658-70.
190. Nguyen, A.P. and D.J. Moore, *Understanding the GTPase Activity of LRRK2: Regulation, Function, and Neurotoxicity.* Adv Neurobiol, 2017. **14**: p. 71-88.
191. Liao, J., et al., *Parkinson disease-associated mutation R1441H in LRRK2 prolongs the "active state" of its GTPase domain.* Proc Natl Acad Sci U S A, 2014. **111**(11): p. 4055-60.
192. Iannotta, L., et al., *Divergent Effects of G2019S and R1441C LRRK2 Mutations on LRRK2 and Rab10 Phosphorylations in Mouse Tissues.* Cells, 2020. **9**(11).
193. Kluss, J.H., et al., *Preclinical modeling of chronic inhibition of the Parkinson's disease associated kinase LRRK2 reveals altered function of the endolysosomal system in vivo.* Mol Neurodegener, 2021. **16**(1): p. 17.
194. Wu, C.X., et al., *Parkinson's disease-associated mutations in the GTPase domain of LRRK2 impair its nucleotide-dependent conformational dynamics.* J Biol Chem, 2019. **294**(15): p. 5907-5913.
195. Huang, X., et al., *The Parkinson's disease-associated mutation N1437H impairs conformational dynamics in the G domain of LRRK2.* FASEB J, 2019. **33**(4): p. 4814-4823.

196. Daniëls, V., et al., *Insight into the mode of action of the LRRK2 Y1699C pathogenic mutant.* J Neurochem, 2011. **116**(2): p. 304-15.
197. Jorgensen, N.D., et al., *The WD40 domain is required for LRRK2 neurotoxicity.* PLoS One, 2009. **4**(12): p. e8463.
198. Piccoli, G., et al., *Leucine-rich repeat kinase 2 binds to neuronal vesicles through protein interactions mediated by its C-terminal WD40 domain.* Mol Cell Biol, 2014. **34**(12): p. 2147-61.
199. Purlyte, E., et al., *Rab29 activation of the Parkinson's disease-associated LRRK2 kinase.* EMBO J, 2018. **37**(1): p. 1-18.
200. Vides, E.G., et al., *A feed-forward pathway drives LRRK2 kinase membrane recruitment and activation.* Elife, 2022. **11**.
201. Antoniou, N., et al., *A motif within the armadillo repeat of Parkinson's-linked LRRK2 interacts with FADD to hijack the extrinsic death pathway.* Sci Rep, 2018. **8**(1): p. 3455.
202. Ding, X. and M.S. Goldberg, *Regulation of LRRK2 stability by the E3 ubiquitin ligase CHIP.* PLoS One, 2009. **4**(6): p. e5949.
203. Polymeropoulos, M.H., et al., *Mutation in the α-Synuclein Gene Identified in Families with Parkinson's Disease.* Science, 1997. **276**(5321): p. 2045-2047.
204. Mergenthaler, P., et al., *Sugar for the brain: the role of glucose in physiological and pathological brain function.* Trends Neurosci, 2013. **36**(10): p. 587-97.
205. Langston, J.W., et al., *Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis.* Science, 1983. **219**(4587): p. 979-80.
206. Tanner, C.M., et al., *Rotenone, paraquat, and Parkinson's disease.* Environ Health Perspect, 2011. **119**(6): p. 866-72.
207. Schapira, A.H., et al., *Mitochondrial complex I deficiency in Parkinson's disease.* J Neurochem, 1990. **54**(3): p. 823-7.
208. Keeney, P.M., et al., *Parkinson's disease brain mitochondrial complex I has oxidatively damaged subunits and is functionally impaired and misassembled.* J Neurosci, 2006. **26**(19): p. 5256-64.
209. Shoffner, J.M., et al., *Mitochondrial oxidative phosphorylation defects in Parkinson's disease.* Ann Neurol, 1991. **30**(3): p. 332-9.
210. Dexter, D.T., et al., *Basal lipid peroxidation in substantia nigra is increased in Parkinson's disease.* J Neurochem, 1989. **52**(2): p. 381-9.
211. Meiser, J., D. Weindl, and K. Hiller, *Complexity of dopamine metabolism.* Cell Commun Signal, 2013. **11**(1): p. 34.
212. Mortiboys, H., et al., *Mitochondrial impairment in patients with Parkinson disease with the G2019S mutation in LRRK2.* Neurology, 2010. **75**(22): p. 2017-20.
213. Li, W.W., et al., *Localization of alpha-synuclein to mitochondria within midbrain of mice.* Neuroreport, 2007. **18**(15): p. 1543-6.
214. Mahul-Mellier, A.L., et al., *The process of Lewy body formation, rather than simply α-synuclein fibrillization, is one of the major drivers of neurodegeneration.* Proc Natl Acad Sci U S A, 2020. **117**(9): p. 4971-4982.

215. Martin, L.J., et al., *Parkinson's disease alpha-synuclein transgenic mice develop neuronal mitochondrial degeneration and cell death.* J Neurosci, 2006. **26**(1): p. 41-50.
216. Chen, R., et al., *Parkinson's disease protein DJ-1 regulates ATP synthase protein components to increase neuronal process outgrowth.* Cell Death Dis, 2019. **10**(6): p. 469.
217. Valente, E.M., et al., *Localization of a novel locus for autosomal recessive early-onset parkinsonism, PARK6, on human chromosome 1p35-p36.* Am J Hum Genet, 2001. **68**(4): p. 895-900.
218. Kitada, T., et al., *Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism.* Nature, 1998. **392**(6676): p. 605-8.
219. Palacino, J.J., et al., *Mitochondrial dysfunction and oxidative damage in parkin-deficient mice.* J Biol Chem, 2004. **279**(18): p. 18614-22.
220. Clark, I.E., et al., *Drosophila pink1 is required for mitochondrial function and interacts genetically with parkin.* Nature, 2006. **441**(7097): p. 1162-6.
221. Park, J., et al., *Mitochondrial dysfunction in Drosophila PINK1 mutants is complemented by parkin.* Nature, 2006. **441**(7097): p. 1157-61.
222. Narendra, D.P., et al., *PINK1 is selectively stabilized on impaired mitochondria to activate Parkin.* PLoS Biol, 2010. **8**(1): p. e1000298.
223. Kondapalli, C., et al., *PINK1 is activated by mitochondrial membrane potential depolarization and stimulates Parkin E3 ligase activity by phosphorylating Serine 65.* Open Biol, 2012. **2**(5): p. 120080.
224. Ordureau, A., et al., *Quantitative proteomics reveal a feedforward mechanism for mitochondrial PARKIN translocation and ubiquitin chain synthesis.* Mol Cell, 2014. **56**(3): p. 360-375.
225. Kane, L.A., et al., *PINK1 phosphorylates ubiquitin to activate Parkin E3 ubiquitin ligase activity.* J Cell Biol, 2014. **205**(2): p. 143-53.
226. Kazlauskaite, A., et al., *Parkin is activated by PINK1-dependent phosphorylation of ubiquitin at Ser65.* Biochem J, 2014. **460**(1): p. 127-39.
227. Koyano, F., et al., *Ubiquitin is phosphorylated by PINK1 to activate parkin.* Nature, 2014. **510**(7503): p. 162-6.
228. Onishi, M., et al., *Molecular mechanisms and physiological functions of mitophagy.* EMBO J, 2021. **40**(3): p. e104705.
229. Gan, Z.Y., et al., *Activation mechanism of PINK1.* Nature, 2022. **602**(7896): p. 328-335.
230. McWilliams, T.G., et al., *Basal Mitophagy Occurs Independently of PINK1 in Mouse Tissues of High Metabolic Demand.* Cell Metab, 2018. **27**(2): p. 439-449.e5.
231. Bi, M., et al., *Expanding the role of proteasome homeostasis in Parkinson's disease: beyond protein breakdown.* Cell Death Dis, 2021. **12**(2): p. 154.
232. McNaught, K.S. and P. Jenner, *Proteasomal function is impaired in substantia nigra in Parkinson's disease.* Neurosci Lett, 2001. **297**(3): p. 191-4.
233. McNaught, K.S., et al., *Altered proteasomal function in sporadic Parkinson's disease.* Exp Neurol, 2003. **179**(1): p. 38-46.

234. Wahl, C., et al., *A comprehensive genetic study of the proteasomal subunit S6 ATPase in German Parkinson's disease patients.* J Neural Transm (Vienna), 2008. **115**(8): p. 1141-8.
235. Beyer, K., M. Domingo-Sàbat, and A. Ariza, *Molecular pathology of Lewy body diseases.* Int J Mol Sci, 2009. **10**(3): p. 724-45.
236. Maroteaux, L., J.T. Campanelli, and R.H. Scheller, *Synuclein: a neuron-specific protein localized to the nucleus and presynaptic nerve terminal.* J Neurosci, 1988. **8**(8): p. 2804-15.
237. Fortin, D.L., et al., *Lipid rafts mediate the synaptic localization of alpha-synuclein.* J Neurosci, 2004. **24**(30): p. 6715-23.
238. Kahle, P.J., et al., *Subcellular localization of wild-type and Parkinson's disease-associated mutant alpha -synuclein in human and transgenic mouse brain.* J Neurosci, 2000. **20**(17): p. 6365-73.
239. Braak, H., et al., *Staging of brain pathology related to sporadic Parkinson's disease.* Neurobiol Aging, 2003. **24**(2): p. 197-211.
240. Kim, S., et al., *Transneuronal Propagation of Pathologic α-Synuclein from the Gut to the Brain Models Parkinson's Disease.* Neuron, 2019. **103**(4): p. 627-641.e7.
241. Van Den Berge, N., et al., *Evidence for bidirectional and trans-synaptic parasympathetic and sympathetic propagation of alpha-synuclein in rats.* Acta Neuropathol, 2019. **138**(4): p. 535-550.
242. Volpicelli-Daley, L.A., K.C. Luk, and V.M. Lee, *Addition of exogenous α-synuclein preformed fibrils to primary neuronal cultures to seed recruitment of endogenous α-synuclein to Lewy body and Lewy neurite-like aggregates.* Nat Protoc, 2014. **9**(9): p. 2135-46.
243. Luk, K.C. and V.M. Lee, *Modeling Lewy pathology propagation in Parkinson's disease.* Parkinsonism Relat Disord, 2014. **20 Suppl 1**(0 1): p. S85-7.
244. Marotta, N.P., et al., *Alpha-synuclein from patient Lewy bodies exhibits distinct pathological activity that can be propagated in vitro.* Acta Neuropathol Commun, 2021. **9**(1): p. 188.
245. Siderowf, A., et al., *Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using α-synuclein seed amplification: a cross-sectional study.* Lancet Neurol, 2023. **22**(5): p. 407-417.
246. Espay, A.J. and M.S. Okun, *Abandoning the Proteinopathy Paradigm in Parkinson Disease.* JAMA Neurol, 2023. **80**(2): p. 123-124.
247. Collier, T.J., et al., *Is Alpha-Synuclein Loss-of-Function a Contributor to Parkinsonian Pathology? Evidence from Non-human Primates.* Front Neurosci, 2016. **10**: p. 12.
248. Gorbatyuk, O.S., et al., *In vivo RNAi-mediated alpha-synuclein silencing induces nigrostriatal degeneration.* Mol Ther, 2010. **18**(8): p. 1450-7.
249. Burré, J., et al., *Alpha-synuclein promotes SNARE-complex assembly in vivo and in vitro.* Science, 2010. **329**(5999): p. 1663-7.
250. Fujiwara, H., et al., *alpha-Synuclein is phosphorylated in synucleinopathy lesions.* Nat Cell Biol, 2002. **4**(2): p. 160-4.

251. Hasegawa, M., et al., *Phosphorylated alpha-synuclein is ubiquitinated in alpha-synucleinopathy lesions.* J Biol Chem, 2002. **277**(50): p. 49071-6.
252. Ramalingam, N., et al., *Dynamic physiological α-synuclein S129 phosphorylation is driven by neuronal activity.* NPJ Parkinsons Dis, 2023. **9**(1): p. 4.
253. Lu, G., et al., *Autophagy in health and disease: From molecular mechanisms to therapeutic target.* MedComm (2020), 2022. **3**(3): p. e150.
254. Anglade, P., et al., *Apoptosis and autophagy in nigral neurons of patients with Parkinson's disease.* Histol Histopathol, 1997. **12**(1): p. 25-31.
255. Alvarez-Erviti, L., et al., *Chaperone-mediated autophagy markers in Parkinson disease brains.* Arch Neurol, 2010. **67**(12): p. 1464-72.
256. Chu, Y., et al., *Alterations in lysosomal and proteasomal markers in Parkinson's disease: relationship to alpha-synuclein inclusions.* Neurobiol Dis, 2009. **35**(3): p. 385-98.
257. Dehay, B., et al., *Pathogenic lysosomal depletion in Parkinson's disease.* J Neurosci, 2010. **30**(37): p. 12535-44.
258. Rocha, E.M., et al., *LRRK2 inhibition prevents endolysosomal deficits seen in human Parkinson's disease.* Neurobiol Dis, 2020. **134**: p. 104626.
259. Smith, L. and A.H.V. Schapira, *Variants and Parkinson Disease: Mechanisms and Treatments.* Cells, 2022. **11**(8).
260. Rocha, E.M., et al., *Progressive decline of glucocerebrosidase in aging and Parkinson's disease.* Ann Clin Transl Neurol, 2015. **2**(4): p. 433-8.
261. Gegg, M.E., et al., *Glucocerebrosidase deficiency in substantia nigra of parkinson disease brains.* Ann Neurol, 2012. **72**(3): p. 455-63.
262. Yahalom, G., et al., *Carriers of both GBA and LRRK2 mutations, compared to carriers of either, in Parkinson's disease: Risk estimates and genotype-phenotype correlations.* Parkinsonism Relat Disord, 2019. **62**: p. 179-184.
263. Omer, N., et al., *A Possible Modifying Effect of the G2019S Mutation in the LRRK2 Gene on GBA Parkinson's Disease.* Mov Disord, 2020. **35**(7): p. 1249-1253.
264. Wandinger-Ness, A. and M. Zerial, *Rab proteins and the compartmentalization of the endosomal system.* Cold Spring Harb Perspect Biol, 2014. **6**(11): p. a022616.
265. Steger, M., et al., *Systematic proteomic analysis of LRRK2-mediated Rab GTPase phosphorylation establishes a connection to ciliogenesis.* Elife, 2017. **6**.
266. MacLeod, D.A., et al., *RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk.* Neuron, 2013. **77**(3): p. 425-39.
267. Boecker, C.A., et al., *Increased LRRK2 kinase activity alters neuronal autophagy by disrupting the axonal transport of autophagosomes.* Curr Biol, 2021. **31**(10): p. 2140-2154.e6.
268. Dou, D., et al., *Regulatory imbalance between LRRK2 kinase, PPM1H phosphatase, and ARF6 GTPase disrupts the axonal transport of autophagosomes.* Cell Rep, 2023. **42**(5): p. 112448.
269. Hetz, C., K. Zhang, and R.J. Kaufman, *Mechanisms, regulation and functions of the unfolded protein response.* Nat Rev Mol Cell Biol, 2020. **21**(8): p. 421-438.
270. Hoozemans, J.J., et al., *Activation of the unfolded protein response in Parkinson's disease.* Biochem Biophys Res Commun, 2007. **354**(3): p. 707-11.

271. Baek, J.H., et al., *Unfolded protein response is activated in Lewy body dementias.* Neuropathol Appl Neurobiol, 2016. **42**(4): p. 352-65.
272. Baek, J.H., et al., *GRP78 Level Is Altered in the Brain, but Not in Plasma or Cerebrospinal Fluid in Parkinson's Disease Patients.* Front Neurosci, 2019. **13**: p. 697.
273. Conn, K.J., et al., *Identification of the protein disulfide isomerase family member PDIp in experimental Parkinson's disease and Lewy body pathology.* Brain Res, 2004. **1022**(1-2): p. 164-72.
274. Uehara, T., et al., *S-nitrosylated protein-disulphide isomerase links protein misfolding to neurodegeneration.* Nature, 2006. **441**(7092): p. 513-7.
275. Sugeno, N., et al., *Serine 129 phosphorylation of alpha-synuclein induces unfolded protein response-mediated cell death.* J Biol Chem, 2008. **283**(34): p. 23179-88.
276. Lee, J.H., et al., *Parkinson's disease-associated LRRK2-G2019S mutant acts through regulation of SERCA activity to control ER stress in astrocytes.* Acta Neuropathol Commun, 2019. **7**(1): p. 68.
277. Chartier-Harlin, M.C., et al., *Translation initiator EIF4G1 mutations in familial Parkinson disease.* Am J Hum Genet, 2011. **89**(3): p. 398-406.
278. Jan, A., et al., *Activity of translation regulator eukaryotic elongation factor-2 kinase is increased in Parkinson disease brain and its inhibition reduces alpha synuclein toxicity.* Acta Neuropathol Commun, 2018. **6**(1): p. 54.
279. Flinkman, D., et al., *Regulators of proteostasis are translationally repressed in fibroblasts from patients with sporadic and LRRK2-G2019S Parkinson's disease.* NPJ Parkinsons Dis, 2023. **9**(1): p. 20.
280. Martin, I., et al., *Ribosomal protein s15 phosphorylation mediates LRRK2 neurodegeneration in Parkinson's disease.* Cell, 2014. **157**(2): p. 472-85.
281. Kim, J.W., et al., *Dysregulated mRNA Translation in the G2019S LRRK2 and LRRK2 Knock-Out Mouse Brains.* eNeuro, 2021. **8**(6).
282. Jang, Y., et al., *Mass Spectrometry-Based Proteomics Analysis of Human Substantia Nigra From Parkinson's Disease Patients Identifies Multiple Pathways Potentially Involved in the Disease.* Mol Cell Proteomics, 2023. **22**(1): p. 100452.
283. McGeer, P.L., et al., *Reactive microglia are positive for HLA-DR in the substantia nigra of Parkinson's and Alzheimer's disease brains.* Neurology, 1988. **38**(8): p. 1285-91.
284. Brochard, V., et al., *Infiltration of CD4+ lymphocytes into the brain contributes to neurodegeneration in a mouse model of Parkinson disease.* J Clin Invest, 2009. **119**(1): p. 182-92.
285. Grozdanov, V., et al., *Inflammatory dysregulation of blood monocytes in Parkinson's disease patients.* Acta Neuropathol, 2014. **128**(5): p. 651-63.
286. Tansey, M.G., et al., *Inflammation and immune dysfunction in Parkinson disease.* Nat Rev Immunol, 2022. **22**(11): p. 657-673.
287. Terkelsen, M.H., et al., *Neuroinflammation and Immune Changes in Prodromal Parkinson's Disease and Other Synucleinopathies.* J Parkinsons Dis, 2022. **12**(s1): p. S149-S163.

288. Muñoz-Delgado, L., et al., *Peripheral inflammatory immune response differs among sporadic and familial Parkinson's disease.* NPJ Parkinsons Dis, 2023. **9**(1): p. 12.
289. Yamada, T., P.L. McGeer, and E.G. McGeer, *Lewy bodies in Parkinson's disease are recognized by antibodies to complement proteins.* Acta Neuropathol, 1992. **84**(1): p. 100-4.
290. Loeffler, D.A., D.M. Camp, and S.B. Conant, *Complement activation in the Parkinson's disease substantia nigra: an immunocytochemical study.* J Neuroinflammation, 2006. **3**: p. 29.
291. Abdi, I.Y., S.S. Ghanem, and O.M. El-Agnaf, *Immune-related biomarkers for Parkinson's disease.* Neurobiol Dis, 2022. **170**: p. 105771.
292. Mackmull, M.T., et al., *Global, in situ analysis of the structural proteome in individuals with Parkinson's disease to identify a new class of biomarker.* Nat Struct Mol Biol, 2022. **29**(10): p. 978-989.
293. Winchester, L., et al., *Identification of a possible proteomic biomarker in Parkinson's disease: discovery and replication in blood, brain and cerebrospinal fluid.* Brain Commun, 2023. **5**(1): p. fcac343.
294. Gregersen, E., et al., *Alpha-synuclein activates the classical complement pathway and mediates complement-dependent cell toxicity.* J Neuroinflammation, 2021. **18**(1): p. 177.
295. Bliederhaeuser, C., et al., *LRRK2 contributes to monocyte dysregulation in Parkinson's disease.* Acta Neuropathol Commun, 2016. **4**(1): p. 123.
296. Cook, D.A., et al., *LRRK2 levels in immune cells are increased in Parkinson's disease.* NPJ Parkinsons Dis, 2017. **3**: p. 11.
297. Lin, J.C., et al., *Association Between Parkinson's Disease and Inflammatory Bowel Disease: a Nationwide Taiwanese Retrospective Cohort Study.* Inflamm Bowel Dis, 2016. **22**(5): p. 1049-55.
298. Qin, L., et al., *Systemic LPS causes chronic neuroinflammation and progressive neurodegeneration.* Glia, 2007. **55**(5): p. 453-62.
299. Kozina, E., et al., *Mutant LRRK2 mediates peripheral and central immune responses leading to neurodegeneration in vivo.* Brain, 2018. **141**(6): p. 1753-1769.
300. Thévenet, J., et al., *Regulation of LRRK2 expression points to a functional role in human monocyte maturation.* PLoS One, 2011. **6**(6): p. e21519.
301. Russo, I., L. Bubacco, and E. Greggio, *LRRK2 as a target for modulating immune system responses.* Neurobiol Dis, 2022. **169**: p. 105724.
302. Visscher, P.M., et al., *10 Years of GWAS Discovery: Biology, Function, and Translation.* Am J Hum Genet, 2017. **101**(1): p. 5-22.
303. Mattick, J.S., et al., *Long non-coding RNAs: definitions, functions, challenges and recommendations.* Nat Rev Mol Cell Biol, 2023. **24**(6): p. 430-447.
304. EDMAN, P., *A method for the determination of amino acid sequence in peptides.* Arch Biochem, 1949. **22**(3): p. 475.
305. Renart, J., J. Reiser, and G.R. Stark, *Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: a method for studying*

antibody specificity and antigen structure. Proc Natl Acad Sci U S A, 1979. **76**(7): p. 3116-20.
306. Beck, M., et al., *The quantitative proteome of a human cell line.* Mol Syst Biol, 2011. **7**: p. 549.
307. Smith, L.M., N.L. Kelleher, and C.f.T.D. Proteomics, *Proteoform: a single term describing protein complexity.* Nat Methods, 2013. **10**(3): p. 186-7.
308. Aebersold, R., et al., *How many human proteoforms are there?* Nat Chem Biol, 2018. **14**(3): p. 206-214.
309. Nickerson, J.L., et al., *Recent advances in top-down proteome sample processing ahead of MS analysis.* Mass Spectrom Rev, 2023. **42**(2): p. 457-495.
310. Melby, J.A., et al., *Novel Strategies to Address the Challenges in Top-Down Proteomics.* J Am Soc Mass Spectrom, 2021. **32**(6): p. 1278-1294.
311. Giles, K., et al., *A Cyclic Ion Mobility-Mass Spectrometry System.* Anal Chem, 2019. **91**(13): p. 8564-8573.
312. Huguet, R., et al., *Proton Transfer Charge Reduction Enables High-Throughput Top-Down Analysis of Large Proteoforms.* Anal Chem, 2019. **91**(24): p. 15732-15739.
313. Coradin, M., et al., *Bullet points to evaluate the performance of the middle-down proteomics workflow for histone modification analysis.* Methods, 2020. **184**: p. 86-92.
314. Scheltema, R.A., et al., *The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer.* Mol Cell Proteomics, 2014. **13**(12): p. 3698-708.
315. Kawashima, Y., et al., *Optimization of Data-Independent Acquisition Mass Spectrometry for Deep and Highly Sensitive Proteomic Analysis.* Int J Mol Sci, 2019. **20**(23).
316. Bekker-Jensen, D.B., et al., *A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients.* Mol Cell Proteomics, 2020. **19**(4): p. 716-729.
317. Wang, Z., et al., *High-throughput proteomics of nanogram-scale samples with Zeno SWATH MS.* Elife, 2022. **11**.
318. Zubarev, R.A., *The challenge of the proteome dynamic range and its implications for in-depth proteomics.* Proteomics, 2013. **13**(5): p. 723-6.
319. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects.* Mol Cell Proteomics, 2002. **1**(11): p. 845-67.
320. Vandermarliere, E., M. Mueller, and L. Martens, *Getting intimate with trypsin, the leading protease in proteomics.* Mass Spectrom Rev, 2013. **32**(6): p. 453-65.
321. Laskay, Ü., et al., *Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down proteomics experiments.* J Proteome Res, 2013. **12**(12): p. 5558-69.
322. Varnavides, G., et al., *In Search of a Universal Method: A Comparative Survey of Bottom-Up Proteomics Sample Preparation Methods.* J Proteome Res, 2022. **21**(10): p. 2397-2411.

323. Ludwig, K.R., M.M. Schroll, and A.B. Hummon, *Comparison of In-Solution, FASP, and S-Trap Based Digestion Methods for Bottom-Up Proteomic Studies.* J Proteome Res, 2018. **17**(7): p. 2480-2490.
324. Choksawangkarn, W., et al., *Comparative study of workflows optimized for in-gel, in-solution, and on-filter proteolysis in the analysis of plasma membrane proteins.* J Proteome Res, 2012. **11**(5): p. 3030-4.
325. Wiśniewski, J.R., P. Ostasiewicz, and M. Mann, *High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers.* J Proteome Res, 2011. **10**(7): p. 3040-9.
326. Klont, F., et al., *Assessment of Sample Preparation Bias in Mass Spectrometry-Based Proteomics.* Anal Chem, 2018. **90**(8): p. 5405-5413.
327. 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.
328. Zhou, J.Y., et al., *Simple sodium dodecyl sulfate-assisted sample preparation method for LC-MS-based proteomics applications.* Anal Chem, 2012. **84**(6): p. 2862-7.
329. Masuda, T., M. Tomita, and Y. Ishihama, *Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis.* J Proteome Res, 2008. **7**(2): p. 731-40.
330. Doellinger, J., et al., *Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-free Protocol for Proteomics Based on Acid Extraction.* Mol Cell Proteomics, 2020. **19**(1): p. 209-222.
331. Shevchenko, A., et al., *In-gel digestion for mass spectrometric characterization of proteins and proteomes.* Nat Protoc, 2006. **1**(6): p. 2856-60.
332. Swart, C., et al., *Hit-Gel: Streamlining in-gel protein digestion for high-throughput proteomics experiments.* Sci Rep, 2018. **8**(1): p. 8582.
333. 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.
334. Hughes, C.S., et al., *Single-pot, solid-phase-enhanced sample preparation for proteomics experiments.* Nat Protoc, 2019. **14**(1): p. 68-85.
335. Zougman, A., et al., *Detergent-Free Simultaneous Sample Preparation Method for Proteomics and Metabolomics.* J Proteome Res, 2020. **19**(7): p. 2838-2844.
336. 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.
337. Wiśniewski, J.R., et al., *Universal sample preparation method for proteome analysis.* Nat Methods, 2009. **6**(5): p. 359-62.
338. Kiick, K.L., et al., *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation.* Proc Natl Acad Sci U S A, 2002. **99**(1): p. 19-24.
339. Beatty, K.E., et al., *Fluorescence visualization of newly synthesized proteins in mammalian cells.* Angew Chem Int Ed Engl, 2006. **45**(44): p. 7364-7.

340. Kolb, H.C., M.G. Finn, and K.B. Sharpless, *Click Chemistry: Diverse Chemical Function from a Few Good Reactions.* Angew Chem Int Ed Engl, 2001. **40**(11): p. 2004-2021.
341. Dieterich, D.C., et al., *Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT).* Proc Natl Acad Sci U S A, 2006. **103**(25): p. 9482-7.
342. Schiapparelli, L.M., et al., *Direct detection of biotinylated proteins by mass spectrometry.* J Proteome Res, 2014. **13**(9): p. 3966-78.
343. McClatchy, D.B., et al., *Pulsed Azidohomoalanine Labeling in Mammals (PALM) Detects Changes in Liver-Specific LKB1 Knockout Mice.* J Proteome Res, 2015. **14**(11): p. 4815-22.
344. Schmidt, E.K., et al., *SUnSET, a nonradioactive method to monitor protein synthesis.* Nat Methods, 2009. **6**(4): p. 275-7.
345. Forester, C.M., et al., *Revealing nascent proteomics in signaling pathways and cell differentiation.* Proc Natl Acad Sci U S A, 2018. **115**(10): p. 2353-2358.
346. NATHANS, D., *PUROMYCIN INHIBITION OF PROTEIN SYNTHESIS: INCORPORATION OF PUROMYCIN INTO PEPTIDE CHAINS.* Proc Natl Acad Sci U S A, 1964. **51**(4): p. 585-92.
347. Liu, J., et al., *Imaging protein synthesis in cells and tissues with an alkyne analog of puromycin.* Proc Natl Acad Sci U S A, 2012. **109**(2): p. 413-8.
348. Ignacio, B.J., et al., *THRONCAT: metabolic labeling of newly synthesized proteins using a bioorthogonal threonine analog.* Nat Commun, 2023. **14**(1): p. 3367.
349. Mahdavi, A., et al., *Engineered Aminoacyl-tRNA Synthetase for Cell-Selective Analysis of Mammalian Protein Synthesis.* J Am Chem Soc, 2016. **138**(13): p. 4278-81.
350. Martínez-Val, A., et al., *Hybrid-DIA: intelligent data acquisition integrates targeted and discovery proteomics to analyze phospho-signaling in single spheroids.* Nat Commun, 2023. **14**(1): p. 3599.
351. 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.
352. Cox, J., et al., *Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ.* Mol Cell Proteomics, 2014. **13**(9): p. 2513-26.
353. Kim, M.S., et al., *A draft map of the human proteome.* Nature, 2014. **509**(7502): p. 575-81.
354. Wilhelm, M., et al., *Mass-spectrometry-based draft of the human proteome.* Nature, 2014. **509**(7502): p. 582-7.
355. Bekker-Jensen, D.B., et al., *An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes.* Cell Syst, 2017. **4**(6): p. 587-599.e4.
356. Wang, D., et al., *A deep proteome and transcriptome abundance atlas of 29 healthy human tissues.* Mol Syst Biol, 2019. **15**(2): p. e8503.

357. Meier, F., et al., *BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes.* Nat Methods, 2018. **15**(6): p. 440-448.
358. Mayer, R.L., et al., *Wide Window Acquisition and AI-based data analysis to reach deep proteome coverage for a wide sample range, including single cell proteomic inputs.* bioRxiv, 2022: p. 2022.09.01.506203.
359. 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.
360. 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.
361. Zeng, D. and S. Li, *Revival of deuterium-labeled reagents for protein quantitation.* Chem Commun (Camb), 2009(23): p. 3369-71.
362. Xiang, F., et al., *N,N-dimethyl leucines as novel isobaric tandem mass tags for quantitative proteomics and peptidomics.* Anal Chem, 2010. **82**(7): p. 2817-25.
363. Ren, Y., et al., *Reagents for Isobaric Labeling Peptides in Quantitative Proteomics.* Anal Chem, 2018. **90**(21): p. 12366-12371.
364. 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.
365. Choe, L., et al., *8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease.* Proteomics, 2007. **7**(20): p. 3651-60.
366. Frost, D.C., Y. Feng, and L. Li, *21-plex DiLeu Isobaric Tags for High-Throughput Quantitative Proteomics.* Anal Chem, 2020. **92**(12): p. 8228-8234.
367. Zecha, J., et al., *TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach.* Mol Cell Proteomics, 2019. **18**(7): p. 1468-1478.
368. Erickson, B.K., et al., *A Strategy to Combine Sample Multiplexing with Targeted Proteomics Assays for High-Throughput Protein Signature Characterization.* Mol Cell, 2017. **65**(2): p. 361-370.
369. Ctortecka, C. and K. Mechtler, *The rise of single-cell proteomics.* Analytical Science Advances, 2021. **2**(3-4): p. 84-94.
370. Budnik, B., et al., *SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation.* Genome Biol, 2018. **19**(1): p. 161.
371. Zhu, Y., et al., *Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS.* Angew Chem Int Ed Engl, 2018. **57**(38): p. 12370-12374.
372. Brenes, A., et al., *Multibatch TMT Reveals False Positives, Batch Effects and Missing Values.* Mol Cell Proteomics, 2019. **18**(10): p. 1967-1980.
373. Yu, S.H., P. Kyriakidou, and J. Cox, *Isobaric Matching between Runs and Novel PSM-Level Normalization in MaxQuant Strongly Improve Reporter Ion-Based Quantification.* J Proteome Res, 2020. **19**(10): p. 3945-3954.
374. Bantscheff, M., et al., *Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer.* Mol Cell Proteomics, 2008. **7**(9): p. 1702-13.

375. Ow, S.Y., et al., *iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly".* J Proteome Res, 2009. **8**(11): p. 5347-55.
376. Karp, N.A., et al., *Addressing accuracy and precision issues in iTRAQ quantitation.* Mol Cell Proteomics, 2010. **9**(9): p. 1885-97.
377. Shirran, S.L. and C.H. Botting, *A comparison of the accuracy of iTRAQ quantification by nLC-ESI MSMS and nLC-MALDI MSMS methods.* J Proteomics, 2010. **73**(7): p. 1391-403.
378. Ting, L., et al., *MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics.* Nat Methods, 2011. **8**(11): p. 937-40.
379. Wenger, C.D., et al., *Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging.* Nat Methods, 2011. **8**(11): p. 933-5.
380. Ogata, K. and Y. Ishihama, *Extending the Separation Space with Trapped Ion Mobility Spectrometry Improves the Accuracy of Isobaric Tag-Based Quantitation in Proteomic LC/MS/MS.* Anal Chem, 2020. **92**(12): p. 8037-8040.
381. Ping, L., et al., *Global quantitative analysis of the human brain proteome in Alzheimer's and Parkinson's Disease.* Sci Data, 2018. **5**: p. 180036.
382. Krieger, J.R., et al., *Evosep One Enables Robust Deep Proteome Coverage Using Tandem Mass Tags while Significantly Reducing Instrument Time.* J Proteome Res, 2019. **18**(5): p. 2346-2353.
383. Orre, L.M., et al., *SubCellBarCode: Proteome-wide Mapping of Protein Localization and Relocalization.* Mol Cell, 2019. **73**(1): p. 166-182.e7.
384. Jiang, L., et al., *A Quantitative Proteome Map of the Human Body.* Cell, 2020. **183**(1): p. 269-283.e19.
385. Zhu, Y., et al., *Discovery of coding regions in the human genome by integrated proteogenomics analysis workflow.* Nat Commun, 2018. **9**(1): p. 903.
386. Bruderer, R., et al., *Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results.* Mol Cell Proteomics, 2017. **16**(12): p. 2296-2309.
387. Tsou, C.C., et al., *DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics.* Nat Methods, 2015. **12**(3): p. 258-64, 7 p following 264.
388. Sinitcyn, P., et al., *MaxDIA enables library-based and library-free data-independent acquisition proteomics.* Nat Biotechnol, 2021. **39**(12): p. 1563-1573.
389. 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.
390. 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.
391. Kong, A.T., et al., *MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics.* Nat Methods, 2017. **14**(5): p. 513-520.
392. Yu, F., et al., *Analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform.* Nat Commun, 2023. **14**(1): p. 4154.

393. Singh, A., *Sensitive protein analysis with plexDIA.* Nat Methods, 2022. **19**(9): p. 1032.
394. van Bentum, M. and M. Selbach, *An Introduction to Advanced Targeted Acquisition Methods.* Mol Cell Proteomics, 2021. **20**: p. 100165.
395. 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.
396. Gallien, S., et al., *Selectivity of LC-MS/MS analysis: implication for proteomics experiments.* J Proteomics, 2013. **81**: p. 148-58.
397. Gallien, S., S.Y. Kim, and B. Domon, *Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM).* Mol Cell Proteomics, 2015. **14**(6): p. 1630-44.
398. Bons, J., et al., *Combining label-free and label-based accurate quantifications with SWATH-MS: Comparison with SRM and PRM for the evaluation of bovine muscle type effects.* Proteomics, 2021. **21**(10): p. e2000214.
399. Escher, C., et al., *Using iRT, a normalized retention time for more targeted measurement of peptides.* Proteomics, 2012. **12**(8): p. 1111-21.
400. Gallien, S., et al., *Technical considerations for large-scale parallel reaction monitoring analysis.* J Proteomics, 2014. **100**: p. 147-59.
401. Yan, W., et al., *Index-ion triggered MS2 ion quantification: a novel proteomics approach for reproducible detection and quantification of targeted proteins in complex mixtures.* Mol Cell Proteomics, 2011. **10**(3): p. M110.005611.
402. Stopfer, L.E., et al., *High-Density, Targeted Monitoring of Tyrosine Phosphorylation Reveals Activated Signaling Networks in Human Tumors.* Cancer Res, 2021. **81**(9): p. 2495-2509.
403. O'Connell, J.D., et al., *Proteome-Wide Evaluation of Two Common Protein Quantification Methods.* J Proteome Res, 2018. **17**(5): p. 1934-1942.
404. Hogrebe, A., et al., *Benchmarking common quantification strategies for large-scale phosphoproteomics.* Nat Commun, 2018. **9**(1): p. 1045.
405. Stepath, M., et al., *Systematic Comparison of Label-Free, SILAC, and TMT Techniques to Study Early Adaption toward Inhibition of EGFR Signaling in the Colorectal Cancer Cell Line DiFi.* J Proteome Res, 2020. **19**(2): p. 926-937.
406. Wang, Z., et al., *Comparison of quantitation methods in proteomics to define relevant toxicological information on AhR activation of HepG2 cells by BaP.* Toxicology, 2021. **448**: p. 152652.
407. Lysiak, A., et al., *Evaluation of open search methods based on theoretical mass spectra comparison.* BMC Bioinformatics, 2021. **22**(Suppl 2): p. 65.
408. 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.
409. Dowell, J.A., et al., *Benchmarking Quantitative Performance in Label-Free Proteomics.* ACS Omega, 2021. **6**(4): p. 2494-2504.
410. Fernández-Costa, C., et al., *Impact of the Identification Strategy on the Reproducibility of the DDA and DIA Results.* J Proteome Res, 2020. **19**(8): p. 3153-3161.

411. Muntel, J., et al., *Comparison of Protein Quantification in a Complex Background by DIA and TMT Workflows with Fixed Instrument Time.* J Proteome Res, 2019. **18**(3): p. 1340-1351.
412. Zhao, J., et al., *Data-independent acquisition boosts quantitative metaproteomics for deep characterization of gut microbiota.* NPJ Biofilms Microbiomes, 2023. **9**(1): p. 4.
413. Heil, L.R., et al., *Evaluating the performance of the Astral mass analyzer for quantitative proteomics using data independent acquisition.* bioRxiv, 2023.
414. Bache, N., et al., *A Novel LC System Embeds Analytes in Pre-formed Gradients for Rapid, Ultra-robust Proteomics.* Mol Cell Proteomics, 2018. **17**(11): p. 2284-2296.
415. Kawashima, Y., et al., *Single-Shot 10K Proteome Approach: Over 10,000 Protein Identifications by Data-Independent Acquisition-Based Single-Shot Proteomics with Ion Mobility Spectrometry.* J Proteome Res, 2022. **21**(6): p. 1418-1427.
416. Verheggen, K., et al., *Anatomy and evolution of database search engines-a central component of mass spectrometry based proteomic workflows.* Mass Spectrom Rev, 2020. **39**(3): p. 292-306.
417. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.
418. Gessulat, S., et al., *Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning.* Nat Methods, 2019. **16**(6): p. 509-518.
419. Välikangas, T., T. Suomi, and L.L. Elo, *A systematic evaluation of normalization methods in quantitative label-free proteomics.* Brief Bioinform, 2018. **19**(1): p. 1-11.
420. Dubois, E., et al., *Assessing normalization methods in mass spectrometry-based proteome profiling of clinical samples.* Biosystems, 2022. **215-216**: p. 104661.
421. Lin, M.H., et al., *Benchmarking differential expression, imputation and quantification methods for proteomics data.* Brief Bioinform, 2022. **23**(3).
422. Jin, L., et al., *A comparative study of evaluating missing value imputation methods in label-free proteomics.* Sci Rep, 2021. **11**(1): p. 1760.
423. Stekhoven, D.J. and P. Bühlmann, *MissForest--non-parametric missing value imputation for mixed-type data.* Bioinformatics, 2012. **28**(1): p. 112-8.
424. Troyanskaya, O., et al., *Missing value estimation methods for DNA microarrays.* Bioinformatics, 2001. **17**(6): p. 520-5.
425. Almeida, J.S., et al., *Normalization and analysis of residual variation in two-dimensional gel electrophoresis for quantitative differential proteomics.* Proteomics, 2005. **5**(5): p. 1242-9.
426. Silva, J.C., et al., *Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition.* Mol Cell Proteomics, 2006. **5**(1): p. 144-56.
427. Schwanhäusser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-42.
428. Belouah, I., et al., *Peptide filtering differently affects the performances of XIC-based quantification methods.* J Proteomics, 2019. **193**: p. 131-141.

429. Shin, J.B., et al., *Molecular architecture of the chick vestibular hair bundle.* Nat Neurosci, 2013. **16**(3): p. 365-74.
430. Chawade, A., E. Alexandersson, and F. Levander, *Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets.* J Proteome Res, 2014. **13**(6): p. 3114-20.
431. Willforss, J., A. Chawade, and F. Levander, *NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis.* J Proteome Res, 2019. **18**(2): p. 732-740.
432. Jones, J., et al., *Tidyproteomics: an open-source R package and data object for quantitative proteomics post analysis and visualization.* BMC Bioinformatics, 2023. **24**(1): p. 239.
433. Choi, M., et al., *MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments.* Bioinformatics, 2014. **30**(17): p. 2524-6.
434. Kohler, D., et al., *MSstats Version 4.0: Statistical Analyses of Quantitative Mass Spectrometry-Based Proteomic Experiments with Chromatography-Based Quantification at Scale.* J Proteome Res, 2023. **22**(5): p. 1466-1482.
435. Goeminne, L.J.E., et al., *MSqRob Takes the Missing Hurdle: Uniting Intensity- and Count-Based Proteomics.* Anal Chem, 2020. **92**(9): p. 6278-6287.
436. Goeminne, L.J., K. Gevaert, and L. Clement, *Peptide-level Robust Ridge Regression Improves Estimation, Sensitivity, and Specificity in Data-dependent Quantitative Label-free Shotgun Proteomics.* Mol Cell Proteomics, 2016. **15**(2): p. 657-68.
437. Yu, F., S.E. Haynes, and A.I. Nesvizhskii, *IonQuant Enables Accurate and Sensitive Label-Free Quantification With FDR-Controlled Match-Between-Runs.* Mol Cell Proteomics, 2021. **20**: p. 100077.
438. Pham, T.V., A.A. Henneman, and C.R. Jimenez, *iq: an R package to estimate relative protein abundances from ion quantification in DIA-MS-based proteomics.* Bioinformatics, 2020. **36**(8): p. 2611-2613.
439. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.
440. Sundararaman, N., et al., *BIRCH: An Automated Workflow for Evaluation, Correction, and Visualization of Batch Effect in Bottom-Up Mass Spectrometry-Based Proteomics Data.* J Proteome Res, 2023. **22**(2): p. 471-481.
441. Plubell, D.L., et al., *Putting Humpty Dumpty Back Together Again: What Does Protein Quantification Mean in Bottom-Up Proteomics?* J Proteome Res, 2022. **21**(4): p. 891-898.
442. Ammar, C., et al., *Accurate Label-Free Quantification by directLFQ to Compare Unlimited Numbers of Proteomes.* Mol Cell Proteomics, 2023. **22**(7): p. 100581.
443. Bludau, I., et al., *Systematic detection of functional proteoform groups from bottom-up proteomic datasets.* Nat Commun, 2021. **12**(1): p. 3810.
444. Kurzawa, N., et al., *Deep thermal profiling for detection of functional proteoform groups.* Nat Chem Biol, 2023.

445. Mylonas, R., et al., *A Database of Accurate Electrophoretic Migration Patterns for Human Proteins.* J Mol Biol, 2023. **435**(4): p. 167933.
446. Liu, Y., A. Beyer, and R. Aebersold, *On the Dependency of Cellular Protein Levels on mRNA Abundance.* Cell, 2016. **165**(3): p. 535-50.
447. Vogel, C., et al., *Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line.* Mol Syst Biol, 2010. **6**: p. 400.
448. Cifani, P., et al., *Hunting for protein markers of hypoxia by combining plasma membrane enrichment with a new approach to membrane protein analysis.* J Proteome Res, 2011. **10**(4): p. 1645-56.
449. Vogel, C. and E.M. Marcotte, *Insights into the regulation of protein abundance from proteomic and transcriptomic analyses.* Nat Rev Genet, 2012. **13**(4): p. 227-32.
450. Edfors, F., et al., *Gene-specific correlation of RNA and protein levels in human cells and tissues.* Mol Syst Biol, 2016. **12**(10): p. 883.
451. Franks, A., E. Airoldi, and N. Slavov, *Post-transcriptional regulation across human tissues.* PLoS Comput Biol, 2017. **13**(5): p. e1005535.
452. Eraslan, B., et al., *Quantification and discovery of sequence determinants of protein-per-mRNA amount in 29 human tissues.* Mol Syst Biol, 2019. **15**(2): p. e8513.
453. Giansanti, P., et al., *Mass spectrometry-based draft of the mouse proteome.* Nat Methods, 2022. **19**(7): p. 803-811.
454. Fortelny, N., et al., *Can we predict protein from mRNA levels?* Nature, 2017. **547**(7664): p. E19-E20.
455. Zhong, W., et al., *Next generation plasma proteome profiling to monitor health and disease.* Nat Commun, 2021. **12**(1): p. 2493.
456. Candia, J., et al., *Assessment of variability in the plasma 7k SomaScan proteomics assay.* Sci Rep, 2022. **12**(1): p. 17147.
457. Katz, D.H., et al., *Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods.* Sci Adv, 2022. **8**(33): p. eabm5164.
458. Raffield, L.M., et al., *Comparison of Proteomic Assessment Methods in Multiple Cohort Studies.* Proteomics, 2020. **20**(12): p. e1900278.
459. Finkernagel, F., et al., *Dual-platform affinity proteomics identifies links between the recurrence of ovarian carcinoma and proteins released into the tumor microenvironment.* Theranostics, 2019. **9**(22): p. 6601-6617.
460. Jarrett, D.E., et al., *A theoretical framework for proteome-scale single-molecule protein identification using multi-affinity protein binding reagents.* bioRxiv, 2021: p. 2021.10.11.463967.
461. Swaminathan, J., et al., *Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures.* Nat Biotechnol, 2018.

Printed by Media-Tryck, Lund 2024 NORDIC SWAN ECOLABEL 3041 0903

Department of Immunotechnology
Faculty of Engineering

ISBN 978-91-8104-046-3