



Technische Universität München



Center of Life and Food Sciences Weihenstephan

Chair of Proteomics and Bioanalytics

Dissertation

Doctor rerum naturalium (Dr. rer. nat.)

Development and Application of Comprehensive Proteomic Methods for the Discovery of Biomarkers of Neurological Conditions.

By

Sebastian T. Berger, M. Sc.

Supervisors:

Prof. Dr. Martin Klingenspor,

Prof. Dr. Bernhard Küster,

Prof. Dr. Hanno Steen

May 2013 to January 2017



Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

Development and Application of Comprehensive Proteomic Methods for the Discovery of Biomarkers of Neurological Conditions

Sebastian Tobias Berger

Vollständiger Abdruck der von der Fakultät

Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:

Prof. Dr. Martin Klingenspor

Prüfer der Dissertation:

1. Prof. Dr. Bernhard Küster
2. Assoc. Prof. Hanno Steen, Ph.D.

Die Dissertation wurde am *18.01.2018* bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am *05.06.2018* angenommen.

(blank page)

Dedicated to my family.

Declaration

I hereby declare, that the present thesis is the result of my own work and includes nothing which is the outcome of collaborative work. Furthermore, I declare that this thesis was not yet part of any form of examination.

No other references or resources than the stated ones have been used. Direct and indirect quotes are noted as such.

Boston, the 01st September 2016

Sebastian T. Berger, M.Sc.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass die vorliegende Arbeit von mir selbst und ohne fremde Hilfe verfasst und noch nicht anderweitig für Prüfungszwecke vorgelegt wurde.

Es wurden keine anderen als die angegebenen Quellen oder Hilfsmittel benutzt. Wörtliche und sinngemäße Zitate sind als solche gekennzeichnet.

Boston, den 01. September 2016

Sebastian T. Berger, M.Sc.

(blank page)

Contents

Declaration	IV
Eidesstattliche Erklärung	IV
Contents	VI
Abstract	XI
Zusammenfassung	XIII
Acknowledgment	XVI
List of tables	XVII
List of figures	XVIII
List of abbreviations	XXI
1. Introduction	1
1.1 Mass spectrometry based proteomics.....	1
1.1.1 Mass Spectrometry.....	3
1.1.1.1 Liquid chromatography based peptide separation	3
1.1.1.2 Ionization Techniques.....	4
1.1.1.3 Mass Analyzer	6
1.1.1.4 Tandem Mass Spectrometry	8
1.1.1.5 Protein Identification	10
1.1.2 Quantitative Mass Spectrometry	11
1.1.2.1 Label-based Quantification.....	11
1.1.2.2 Label-free Quantification	13
1.2 Biomarker	16
1.2.1 Types of biomarker.....	17
1.2.2 Biomarker Discovery	18
1.3 Neurological Injuries	20
1.3.1 Acquired Brain Injury	21
1.3.2 Neurodegenerative Disorders	22
2. MStern Blotting.....	25
2.1 Introduction.....	25
2.2 Experimental Procedures	26
2.2.1 Cell Culture	26

2.2.2	Urine Collection for Ovarian Cyst Biomarker Study	26
2.2.3	Protein Concentration Determination	26
2.2.4	MStern Blot.....	26
2.2.5	Filter assisted sample preparation	27
2.2.6	LC-MS/MS Analysis.....	27
2.2.7	Data Analysis.....	28
2.3	Results and Discussion	30
2.3.1	FASP vs. MStern Blot.....	30
2.3.2	Performance of MStern Blot	31
2.3.3	Detecting Biases in Proteins Identified in Method-Specific Samples.....	33
2.3.4	Physico-Chemical Properties	33
2.3.5	Protein quantification	35
2.3.6	Ovarian Cyst Biomarker Study	37
2.3.7	Digestion Optimization with Trifluoroethanol.....	38
2.4	Conclusion and Perspective.....	40
2.5	Acknowledgments and Contributions.....	41
3.	Urinary Biomarker Discovery for mTBI.....	42
3.1	Introduction.....	42
3.2	Experimental Procedures	44
3.2.1	Urine Sample Collection	44
3.2.2	MStern Blot.....	44
3.2.3	LC-MS/MS Analysis.....	45
3.2.3.1	Information Dependent Acquisition (IDA) for Ion Library Generation.....	45
3.2.3.2	Data-dependent Acquisition for the Biomarker Qualification Ion Library	45
3.2.3.3	Sequential Windowed Acquisition of All Theoretical MS (SWATH)	46
3.2.3.4	DIA Analysis for Biomarker Qualification.....	46
3.2.4	Data Analysis.....	47
3.2.4.1	DIA Analysis for Biomarker Discovery.....	47
3.2.4.2	DIA Analysis for Biomarker Qualification.....	48
3.3	Results and Discussion	49
3.3.1	Sample cohorts.....	49

3.3.2	Biomarker Discovery	50
3.3.2.1	Ion library and Identifications.....	50
3.3.2.2	SWATH-based Protein Biomarker Discovery.....	50
3.3.3	Biomarker Qualification	55
3.3.3.1	Ion library and protein identifications.....	55
3.3.3.2	Protein quantifications	56
3.3.3.3	Biomarker Qualification of TNFR16, NGAL and RBP4	58
3.3.4	Secondary mTBI Biomarker Discovery	59
3.3.5	Urinary Biomarker Discovery for Intracranial Bleed	61
3.4	Conclusion and Perspective.....	64
3.5	Acknowledgments and Contribution.....	64
4.	Exosomal Biomarker Discovery for mTBI in Urine	65
4.1	Introduction.....	65
4.2	Experimental Procedures	66
4.2.1	Urine Sample Collection	66
4.2.2	Urinary Exosome Extraction	66
4.2.3	Filter-assisted sample preparation (FASP)	66
4.2.4	Graphite Carbon Peptide Fractionation.....	67
4.2.4	LC-MS/MS Analysis.....	67
4.2.4.1	Data-dependent Acquisition for a fractionated ion library (fLib).....	67
4.2.4.2	Data-independent Acquisition for quantitative analysis.....	68
4.2.5	Data Analysis.....	68
4.2.5.1	DIA Analysis	68
4.2.5.2	Statistical Analysis.....	69
4.3	Results and Discussion	70
4.3.1	Sample Cohorts.....	70
4.3.2	Ion Library and Protein Identifications	70
4.3.3	mTBI Biomarker Discovery in the Urinary Exosome Proteome	71
4.3.4	Extended Sample Cohort for Biomarker Qualification.....	75
4.3.5	Intracranial-bleed Biomarker Discovery in Urinary Exosomes.....	76
4.3.5.1	Analysis for involvement in disease or functional processes	78

4.4	Perspective	79
4.5	Acknowledgments and Contributions	80
5.	Terminomal Biomarker Discovery for mTBI	81
5.1	Introduction	81
5.2	Experimental Procedures	82
5.2.1	Identification of semi-specific peptides	82
5.2.2	Peptide-based Quantification	82
5.2.3	Statistical Analysis	82
5.3	Results	83
5.3.1	mTBI-based analysis	83
5.3.2	Intracranial bleed focused analysis	85
5.4	Discussion	87
5.5	Acknowledgments and Contributions	88
6.	CSF-based Biomarker Discovery for NDs	89
6.1	Introduction	89
6.2	Experimental Procedure	90
6.2.1	CSF sample collection	90
6.2.2	MStern Blot	90
6.2.3	LC-MS/MS analysis	90
6.2.3.1	Ion library generation for data-independent acquisition	90
6.2.3.2	Quantitative data acquisition (DIA)	91
6.2.4	DIA data analysis	91
6.2.5	Statistical Analysis	92
6.3	Results and Discussion	93
6.3.1	Sample Cohort	93
6.3.2	Protein Quantification and Ion Library	93
6.3.3	PSP-centered Biomarker Discovery	94
6.3.3.1	Discovery of Diagnostic Biomarkers for PSP	94
6.3.3.2	Discovery of Stratification Biomarker for PSP	97
6.3.4	PD-centered Biomarker Discovery	98
6.3.4.1	Discovery of Diagnostic Biomarkers	98

6.3.4.2	Discovery of Stratification Biomarker Candidates	99
6.3.5	AD-centered Biomarker Discovery	101
6.3.5.1	Discovery of Diagnostic Biomarker Candidates	101
6.3.5.2	Discovery of Stratification Biomarker Candidates	102
6.3.6	CBD-centered Biomarker Discovery	105
6.3.6.1	Discovery of Diagnostic Biomarker Candidates	105
6.3.6.2	Discovery of Stratification Biomarker Candidates	106
6.4	Conclusion and Perspective	109
6.5	Acknowledgements and Contributions	110
7.	References	111
8.	Appendix	128
8.1	MStern Blot	128
8.2	Urinary Biomarker Discovery for mTBI	130
8.2.1	PCA Plots of SWATH-based Quantification	130
8.2.2	PCA Plots of the secondary discovery study	132
8.2.3	Prospective mTBI biomarker candidates	134
8.2.4	PCA Plots of the ICB-based biomarker discovery	135
8.3	Exosomal Biomarker Discovery for mTBI	137
8.3.1	Initial Biomarker Discovery Cohort	137
8.3.2	Extended Sample Cohort for Biomarker Qualification	140
8.3.3	Disease and Function Analysis	141
8.4	Terminomal Biomarker Discovery	142
8.5	Discovery of Biomarkers for NDs in CSF	144
8.5.1	PSP-centered Biomarker Discovery	145
8.5.2	PD-centered Biomarker Discovery	146
8.5.3	AD-centered Biomarker Discovery	146
8.5.4	CBD-centered Biomarker Discovery	147
9.	Curriculum Vitae of the Author	150

Abstract

The nervous system of the human body represents the main control network and thus needs to supervise numerous body functions. Since this system must react to fine external and internal stimuli, already marginal changes within this sensitive network can cause major interruptions, damages or even loss of function. As of today, there are more than 600 different diseases and pathological processes are currently described in literature, which are directly related to nervous system disruption as e.g. genetic disorders (Huntington's disease, muscular dystrophy, etc.), neuro-degenerative diseases (Alzheimer's disease, Parkinson's disease, etc.), injuries of the spinal cord and brain (traumatic brain injury), cancer (brain tumors) or infections (e.g. meningitis). Normally, with the occurrence of symptoms, the underlying disease has already greatly progressed, so that progress inversion or even healing represents a "Mission Impossible". Today, physicians are still very limited with their treatment opportunities, so that as only action the tie or slow-down of the pathological process remains. Nonetheless, once the disease has vastly progressed, even the tie of the underlying process is not enough to help a patient. The key part in these scenarios is the earliest possible detection of the onset of such a disease to prevent a manifestation of such disease.

In this thesis, the scope lays within the development and application of proteomic methods for the discovery of novel biomarkers of neurological conditions. The discovery of novel biomarker molecules normally requests the parallel analysis of many samples in only little time. At best, those samples are also prepared simultaneously. Conventional proteomic sample preparation methods make use of the 96-well plate format to allow for the simultaneous preparation of 96 samples or multiples there-of, at a time. Nonetheless, these methods require several hours to be completed. Here, a novel sample-preparation method was developed and further optimized for specialized application with human urine, cerebrospinal fluid or plasma. This method specifically exploits the naturally occurring hydrophobic interactions between a protein's inner core and a polyvinylidene fluoride membrane, as well as the larger pores of PVDF, which in turn on one hand allows for sample digestion and clean-up directly on the membrane and on the other hand allows for the omittance of prolonging centrifugation steps. Thus, several hours of preparation time can be saved compared to the widely adopted and applied "filter-aided sample preparation" method.

In particular, the aforementioned developed method is employed to analyze the urine of young adolescent after sustaining a concussion. In this context, a concussion is considered a mild-traumatic brain injury which can, especially after multiple occurrences, lead to an inevitable neuro-degenerative disease. Yet, there are no objective measures available that allow physicians to directly diagnose a concussion. Hence, the urinary proteome was specifically investigated within this scope, especially in the regard of discovering urinary biomarker candidates for objectively diagnosing a concussion. Additionally, the associated sample cohort featured dedicated positive controls, suffering a severe traumatic brain injury, which are represented by patients suffering from an intracranial bleed. Later on these samples allowed the initiation of a discovery step for urinary biomarker candidates for intracranial bleeds. However, identified potential candidates, in

the mild-traumatic brain injury as well as in the severe traumatic brain injury study, could yet not be qualified in an enhanced follow-up study.

Besides the investigation of the urinary proteome, also the proteome of urinary exosomes was investigated in the context of mild-traumatic brain injury. Exosomes are small extracellular vesicles (< 100 nm) which are secreted into urine. In particular, exosomes can also feature a signaling function and can transmit messages from one organ to another. Additionally, the urinary proteome consists to a certain extent out of the proteins stored in exosomes. Especially in this context, it was postulated that these nanovesicles can contain valuable information stored as protein pattern, which led consequently led to the investigation via exosome enrichment and proteomic analysis. In the here described analysis, certain proteins in the exosome proteome were found that allowed a certain prediction. In addition, these proteins could directly be linked to various neurological processes. However, the featured specificity and selectivity might not satisfy nowadays clinical standards.

Urine is considered the waste-fluid of the human body. Its main function lays within the evacuation of unwanted metabolites, proteins or other products. Nevertheless, urine is first stored for several hours, in the bladder, before its evacuation through the ureter. During this time or at any time of a proteins' lifetime, endogenous proteases can attack and cut the proteins' backbone. Here, the hypothesis that reoccurrences of specific non-tryptic cleavages within a proteins backbone can trace back to targeted degradation or deactivation processes in the context of mild-traumatic and severe-traumatic brain injuries, was investigated. In the investigated sample cohort 22 peptides for mild-traumatic and 20 peptides for intracranial bleed were discovered. However, a direct linkage between the occurrence of such a specific non-tryptic peptide and the diagnosis of one of the mentioned neurological condition could not be established.

Lastly, cerebrospinal fluid of patients with various neurodegenerative diseases (e.g. Alzheimer's Disease, Parkinson's Disease, or progressive supranuclear palsy) were investigated in the regard of the discovery of disease specific diagnostic or stratification biomarker candidates. In this case, cerebrospinal fluid features special importance as it is produced by ventricles in the brain and circulates within the skull and spinal cord. Furthermore, cerebrospinal fluid can be sampled via a semi-invasive procedure (lumbar puncture). Due to its adjacencies to the brain, this body fluid is considered a valuable resource for the discovery of neurodegenerative diseases. The qualitative and quantitative analysis of the cerebrospinal fluid proteome in this thesis did reveal certain relations between proteins and the underlying diagnosis, which also allows for a certain degree of stratification. Even though, the calculated probability of entering clinical trials for potential candidates in this thesis is low, the discovered biomarker candidate molecules for Alzheimer's disease, Parkinson's disease, progressive supranuclear palsy and corticobasal degeneration can provide insightful information for either reconfirmation of previously identified biomarkers or disease specific processes. Furthermore, these findings might not only be a proof of concept but might provide the gateway for larger scaled studies.

Zusammenfassung

Das menschliche Nervensystem stellt das Hauptkontrollnetzwerk dar und muss aus diesem Grund sehr viele Körperfunktionen überwachen. Da dieses System auf jegliche noch so feinen, externen, wie auch internen Stimuli reagieren muss, können bereits geringe Veränderungen an diesem sensitiven Netzwerk wesentliche Einschränkungen, Beschädigungen oder sogar bis hin zum Verlust der Kontrollfunktion, hervorrufen. Nach heutigem Stand sind mehr als 600 verschiedene Krankheiten und pathologische Prozesse in der aktuellen Literatur beschrieben, welche im Besonderen direkt mit der Zerrüttung des Nervensystems assoziiert sind, so z.B. genetische Erkrankungen (Corea Huntington, muskuläre Dystrophie, etc.), neurodegenerative Krankheiten (z.B. Morbus Alzheimer, Morbus Parkinson, etc.), Verletzungen des Rückenmarks und / oder Gehirns (traumatische Gehirnverletzung, intrakranielle Blutung), Krebs (Gehirn-tumor) oder Infektionen (z.B. Meningitis). Jedoch ist die zugrundeliegende Erkrankung mit dem Auftreten erster Symptome bereits so weit vorangeschritten, sodass der Krankheitsfortschritt nicht umgekehrt, geschweige denn geheilt werden kann. Heutzutage stehen nur geringe Behandlungsoptionen den Ärzten zur Verfügung, so, dass die einzige Möglichkeit darin besteht, den pathologischen Prozess zu stoppen oder gar nur zu verlangsamen. Einst, dass die Krankheit sehr weit vorangeschritten ist, reicht es allerdings meist nicht immer aus nur den Prozess zu stoppen, um einen Patienten zu helfen. Das Schlüsselereignis dieses Szenarios ist die früh-möglichste Erkennung der Krankheit, sodass eine Manifestation des pathologischen Prozesses und somit der Krankheit verhindert werden kann.

Der Fokus dieser Arbeit liegt auf der Entwicklung und Anwendung von proteomischen Methoden zur Entdeckung von neuen Biomarkern von neurologischen Erkrankungen. Besonders die Entdeckung von neuartigen Biomarkermolekülen erfordert normalerweise die parallele Analyse von vielen biologischen Proben in kurzer Zeit. Am Besten werden alle benötigten Proben simultan und somit unter gleichen Bedingungen präpariert. Gerade in diesem Fall behelfen sich konventionelle proteomische Präparationstechniken des 96-well Platten Formats, welches es erlaubt 96 Proben, oder ein Vielfaches davon, simultan zu präparieren. Nichtsdestotrotz benötigen diese Methoden mehrere Stunden zum Durchlaufen, da sie mehrfach Zentrifugationsschritte benötigen, die umgekehrt mehrere Stunden in Anspruch nehmen. Diese Thematik befasst sich im Besonderen mit der Entwicklung sowie Optimierung einer neuartigen Probenaufbereitungsmethode für die spezielle Anwendbarkeit auf humanen Urin, Rückenmarksflüssigkeit oder Blut-Plasma. Die genannte Methode nutzt speziell die natürlich vorhandenen hydrophoben Wechselwirkungen zwischen dem inneren Kern eines denaturierten Proteins und der hydrophoben Oberfläche wie z.B.: polyvinylidene Fluoride (PVDF) aus, um diese direkt an der Membranoberfläche aufzureinigen und tryptisch zu verdauen. Im Besonderen kamen hier die sehr großen Poren der PVDF Membran zur Geltung, die es erlauben einen Flüssigkeitstransfer durchzuführen, ohne jedoch zu zentrifugieren. Hierbei wird das zentrifugieren gänzlich unnötig, was wiederum es erlaubt, mehrere Stunden an Präparationszeit einzusparen – besonders im Vergleich der weithin verbreiteten und bekannten „filter-aided sample preparation“ Methode (FASP).

Im Weiteren wird diese oben genannte, neuentwickelte Methode verwendet, zur Analyse des Urins junger Erwachsener, nach Erleiden einer Gehirnerschütterung. In diesem Zusammenhang wird die Gehirnerschütterung als milde Form der traumatischen Gehirnverletzung gesehen, welche besonders nach häufigem Auftreten, unausweichlich zu einer Erkrankung an einer neurodegenerativen Krankheit führt. Bis heute gibt es allerdings noch keine objektive Bemessungsgrundlage, die es Ärzten erlaubt direkt und unzweifelhaft eine Gehirnerschütterung zu diagnostizieren. Daher wurde in dieser Arbeit spezifisch das Urinproteom untersucht im Besonderen Hinblick auf die Entdeckung neuer Biomarkern zur objektiven Diagnose einer Gehirnerschütterung – im Besten Fall durch einen Urintest. Hierbei wurde besonderen Wert auf das Design der Probenkohorte gelegt und ein Teil „positiver“ Proben, in Form von Urin von Patienten mit intrakranieller Blutung, mit einbezogen. Im weiteren Verlauf der Analyse erlaubte es die Menge an Proben sowohl die Untersuchung zur Entdeckung von Biomarkern, speziell für intrakranielle Blutungen. Dennoch konnten die gefundenen Kandidatenmoleküle sowohl für die milde Form der traumatischen als auch der schweren Form der traumatischen Gehirnverletzung nicht weiter, wie auch in einer erweiterten Nachverfolgungsanalyse, qualifiziert werden.

Neben der Untersuchung des Proteoms im Urin von Patienten, die an einer Gehirnerschütterung erkrankt sind, wurde auch gesondert das Proteom der urinösen Exosomen untersucht. Exosomen sind winzige extrazelluläre Partikel (< 100 nm), welche speziell in den Urin sekretiert werden. Im Besonderen können Exosomen eine signalgebende Funktion übernehmen und ermöglichen es daher ein Signal von einem Organ zum anderen zu transportieren. Zudem besteht das urinäres Proteom zu einem bestimmten Prozentsatz aus den Proteinen, die in diesen Exosomen gespeichert sind. Gerade in diesem Zusammenhang wurde postuliert, dass diese Nanovesikel einen bestimmten Informationsgehalt in einem Proteinnormenmuster speichern. Daher wurde bei dieser Analyse die Exosomen spezifisch angereichert und proteomisch analysiert. In der hier beschriebenen Untersuchung wurden gewisse Proteine, bzw. Proteinnormenmuster identifiziert, die eine gewisse Prognose erlauben. Zusätzlich ist es möglich an Hand dieser Muster gewisse neurologische Prozesse zuzuordnen. Allerdings entspricht die hier gesehene Selektivität und Spezifität nicht den heute geforderten klinischen Standards.

Urin wird als solches als das Abwasser des menschlichen Körpers bezeichnet. Die Hauptaufgabe liegt darin, ungewollte Metaboliten, Proteine und andere Produkte aus dem Körper zu transportieren. Nichtsdestotrotz wird der Urin erst in der Blase für mehrere Stunden gespeichert, bis er über den Harnleiter abtransportiert wird. Die entsprechenden Umgebungsbedingungen stellen beim genaueren Hinsehen die perfekten Bedingungen für endogene Proteasen dar. Daher kann es in dieser Zeit- oder auch zu jeder Zeit während der Lebensdauer eines Proteins dazu kommen von einer solchen Protease angegriffen zu werden. Dies führt nun unweigerlich zu einem entsprechenden Schnitt im Backbone. Hier wurde die Hypothese, dass das Wiederauftreten von spezifischen nicht-tryptischen Schnittstellen innerhalb des Proteinbackbones ein Hinweis auf eine gerichtete Degeneration oder Deaktivierung des Proteins im Kontext der milden sowie schweren traumatischen Gehirnverletzung gibt. Die Analyse des Urins der eingeschriebenen Patienten führte zur Identifikation von 22 solcher Schnittstellenhypothesen für die mild-traumatische und 20

Schnittstellhypothesen für die intrakranielle Blutung, die einen speziellen Hinweis auf das Vorhandensein der Erkrankung liefert. Jedoch konnte eine direkte Verknüpfung zwischen dem Auftreten der Schnittstelle und der jeweiligen Diagnose nicht hergestellt werden.

Zuletzt wurde die Rückenmarksflüssigkeit von Patienten mit verschiedenen neurodegenerativen Erkrankungen (z.B. Morbus Alzheimer, Morbus Parkinson oder Progressive supranukleäre Blickparese) im Hinblick auf die Entdeckung diagnostischer oder stratifizierenden krankheitsspezifischen Biomarkerkandidaten. In diesem Fall stellt die Rückenmarksflüssigkeit eine zentrale Probenart dar, da es von den Ventrikeln im Gehirn gebildet wird und innerhalb des Schädels und dem Rückenmark zirkuliert. Zudem kann diese Körperflüssigkeit vergleichsweise einfach und semi-invasiv (Lumbalpunktion) entnommen werden. Durch die direkte Nachbarschaft zum Gehirngewebe stelle diese Flüssigkeit die zentrale Rolle als Probe zur Entdeckung von Biomarkern neurodegenerativen Erkrankungen dar, da Gehirngewebe nicht zur weiteren Analyse entnommen werden muss. Die qualitative sowie quantitative Analyse des Rückenmarksflüssigkeitsproteom in dieser Arbeit, erlaubte die Entdeckung von bestimmten Beziehungen zwischen dem Auftreten verschiedener Proteinkonzentrationen und der zugrundeliegenden Erkrankung, welche umgekehrt auch einen gewissen Grad an Stratifizierung erlaubt. Obwohl die berechnete Wahrscheinlichkeit zum Eintritt der in dieser Arbeit gefundenen Biomarkerkandidaten für Morbus Alzheimer, Morbus Parkinson, Progressive supranukleäre Blickparese und Kortikobasale Degeneration wertvolle Informationen entweder zur Rekonformierung bereits gefundener Biomarker kiefen oder Hinweise zu Krankheitsbildern geben. Des Weiteren, mögen die hier beschriebenen Ergebnisse nicht nur ein Beweis der Methodenfunktionalität sein, sondern auch die Möglichkeit bieten, größer angelegte Studien anzuregen.

Acknowledgment

Firstly, I would like to express my sincere gratitude to my advisors Prof Dr. Bernhard Küster and Prof Dr. Hanno Steen for their continuous support of my Ph.D. study and related research, for the patience, motivation, and immense knowledge. Their guidance helped me, at all levels and types of research and writing this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisors, I would like to thank Prof. Dr. Judith Steen, Prof. Dr. Hanno Steen and Dr. Jan Muntel as well as Dr. Kevin Broadbelt, for their insightful comments and encouragement, but also for the hard questions, which incited me to widen my research from various perspectives. My sincere thanks also go to Prof Dr. Hanno Steen and Prof. Dr. Judith Steen, who provided me an opportunity to join their team as intern, and who gave access to the laboratory and research facilities. Without their precious support, it would not be possible to conduct this research.

I thank my fellow labmates: Dr. Kevin Broadbelt, Dr. Nerea Cuevas Polo, Dr. Michaela Helmel, Dr. Jan Muntel Saima Ahmed and John K. Sauld; for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last four years. I am grateful to Dr. Joao A. Paulo for enlightening me the first glance of research.

Also, a very special thank you to my wife, who always supported me in this enduring endeavor with all her heart. She never resigned, even in difficult situations and was always supportive in every moment.

Last but not the least, I would like to thank my family, for supporting me spiritually throughout writing this thesis and my life abroad, in general.

List of tables

Table 1.1: Exemplary list of appropriate matrices for MALDI ionization	5
Table 1.2: Exemplary biomarkers used in clinical settings	16
Table 1.3: Exemplary list of symptoms for traumatic brain injuries	22
Table 1.4: Chart of in ND involved mechanisms	23
Table 1.5: Exemplary list of neurodegenerative disease classification	24
Table 2.1: Number of identified proteins per applied organic content composition	38
Table 3.1: Vendor specific entitelements for acquisition modes	43
Table 3.2: Variable extraction windows used in DIA experiments	45
Table 3.3: Observed proteins with significantly different protein abundance	50
Table 3.4: List of observed p-values in the extended cohort	57
Table 4.1: Extraction windows applied for DIA analysis	67
Table 4.2: List of filtered proteins with significant abundance changes	72
Table 4.3: List of protein combinations and resulting AUROC	73
Table 4.4: List of dual significant protein groups	74
Table 4.5: List of protein groups observed as significantly different in both cohorts	75
Table 4.6: List of calculated AUROC values	76
Table 6.1: Variable extraction windows used in DIA experiments	90
Table 6.2: List of proteins observed with significant abundance changes	97
Table 6.3: List of the identified seven stratification biomarker candidates	103
Table 6.4: List of the top three biomarker candidates for CBD	105
Table 6.5: List of the identified seven significant proteins	107
Table 8.2.1: List of identified proteins with significant abundance changes	133
Table 8.3.1: List of associated "Diseases and Functions"	140
Table 8.4.1: Terminomal peptides of interest for diagnosing mTBI	141
Table 8.4.2: Terminomal peptides of interest for diagnosing ICB	142

List of figures

Figure 1.1: Proteomic Workflows	1
Figure 1.2: Interaction Triangle	3
Figure 1.3: van Deemter plot	3
Figure 1.4: Schematic illustration of electrospray ionization	5
Figure 1.5: Schematic illustration of time-of-flight mass analyzer	6
Figure 1.6: The Orbitrap mass analyzer	8
Figure 1.7: MS-based peptide backbone fragmentation	9
Figure 1.8: Quantitative proteomic workflows	11
Figure 1.9: Data-Independent acquisition	15
Figure 1.10: Illustration of “prognostic” and “predictive” biomarkers	17
Figure 1.11: Illustration of the proteomic biomarker discovery workflow	19
Figure 1.12: The brain and its functional areas	20
Figure 1.13: Schematic explanation of “Acquired Brain Injury” (ABI)	21
Figure 2.1: FASP vs. MStern Blot	29
Figure 2.2: Performance Comparison MStern Blot vs. FASP	31
Figure 2.3: Comparison of the Properties of Identified Proteins	33
Figure 2.4: Physical-chemical Properties	34
Figure 2.5: Correlation of FASP- and MStern Blotting-based Protein Quantification	35
Figure 2.6: Statistical Analysis of Serpin B3 and 4 iBAQ intensities in urine	36
Figure 2.7: Impact of TFE on CSF Protein Identification	37
Figure 3.1: Causes of Traumatic Brain Injury	41
Figure 3.2: Q-Q-Plots for samples analyzed by SWATH	49
Figure 3.3: Principal Component Analysis for all quantified proteins	51
Figure 3.4: PCA Bi-Plot (PC1 and PC3) for significantly different abundant proteins	52
Figure 3.5: Evaluation of possible biomarker candidates	53
Figure 3.6: Venn diagram of used ion libraries	55
Figure 3.7: Q-Q-Plots for all 50 DIA analyzed samples	56
Figure 3.8: ROC curve analysis for identified biomarker candidates	57
Figure 3.9: Venn diagram for identified candidates	58
Figure 3.10: 3D PCA plot for the identified 26 significant protein groups	59
Figure 3.11: ROC of biomarker candidate panel	59
Figure 3.12: 3-way Venn diagram for all comparisons	60
Figure 3.13: 3D PCA plot for all 75 significant protein groups	61
Figure 3.14: ROC curve for ICB diagnosing efficiency analysis	62
Figure 4.1: Ion Library Comparison	69
Figure 4.2: GO Annotation for both ion libraries	70
Figure 4.3: Extend of missing values	71
Figure 4.4: PCA Plot of proteins with significantly different protein abundance	72

Figure 4.5: Statistical Analysis of Q9Y2A7 and Q96F07	76
Figure 4.6: Screenshot of Ingenuity® IPA® summary page	77
Figure 5.1: 3D PCA plot for significant observed terminomal peptides	82
Figure 5.2: Supervised PCA Bi-Plot for mTBI biomarker discovery	83
Figure 5.3: 3D PCA plot for the discovery of ICB specific clustering	84
Figure 5.4: PCA Bi-Plot of significantly observed terminomal peptides in ICB	85
Figure 6.1: PCA Bi-Plot for significantly observed proteins (Controls vs. PSP)	93
Figure 6.2: ROC curve for disease biomarker	94
Figure 6.3: ROC curve for healthy biomarker candidates	94
Figure 6.4: PCA plot of linear combined biomarker panel	95
Figure 6.5: PCA plot for stratification	96
Figure 6.6: PCA plot for diagnosing PD in CSF	97
Figure 6.7: 2D scatter plot for PD and Controls	98
Figure 6.8: PCA Bi-Plot for biomarker candidates for PD stratification	99
Figure 6.9: 2D scatter plot for AD diagnosis	100
Figure 6.10: PCA plot for AD stratification	101
Figure 6.11: Extended PCA Bi-Plot for AD stratification	102
Figure 6.12: Supervised PCA plot for CBD biomarker discovery	104
Figure 6.13: Supervised PCA plot for the discovery of stratification biomarker candidates ...	106
Figure 8.1.1: Physio-chemical Properties (CSF)	127
Figure 8.1.2: Physio-chemical Properties (Urine)	128
Figure 8.2.1: Unsupervised mTBI PCA Plot (1)	129
Figure 8.2.2: Unsupervised mTBI PCA Plot (2)	129
Figure 8.2.3: Unsupervised mTBI PCA Plot (3)	130
Figure 8.2.4: Unsupervised mTBI PCA Plot for the secondary discovery (1)	131
Figure 8.2.5: Unsupervised mTBI PCA Plot for the secondary discovery (2)	131
Figure 8.2.6: Unsupervised mTBI PCA Plot for the secondary discovery (3)	132
Figure 8.2.7: Unsupervised ICB PCA Plot (1)	134
Figure 8.2.8: Unsupervised ICB PCA Plot (2)	134
Figure 8.2.9: Supervised PCA Plot for ICB biomarker discovery	135
Figure 8.3.1: Unsupervised PCA Plot sample clustering (1)	136
Figure 8.3.2: Unsupervised PCA Plot sample clustering (2)	136
Figure 8.3.3: Semi-Supervised PCA Plot (1)	137
Figure 8.3.4: Semi-Supervised PCA Plot (2)	137
Figure 8.3.5: Q-Q-Plots for exosomal proteome sample data	138
Figure 8.3.6: PCA Bi-Plot for filtered exosomal proteins	138
Figure 8.3.7: Q-Q-Plots of samples from the extended qualification sample cohort	139
Figure 8.5.1: Q-Q-Plots for CSF samples analyzed by DIA	143
Figure 8.5.2: PCA Plot of all proteins for PSP and Control Samples	144
Figure 8.5.3: PCA Bi-Plot for PSP sample stratification	144

Figure 8.5.4: PCA Bi-Plot for PD biomarker discovery 145

Figure 8.5.5: PCA Bi-Plot for AD stratification 145

Figure 8.5.6: Unsupervised PCA Plot for CBD biomarker discovery 146

Figure 8.5.7: Supervised PCA Bi-Plot for CBD biomarker discovery 146

Figure 8.5.8: Stratification PCA Bi-Plot CBD vs NDs (1) 147

Figure 8.5.9: Stratification PCA Bi-Plot CBD vs NDs (2) 147

Figure 8.5.10: Stratification PCA Bi-Plot CBD vs NDs (3) 148

List of abbreviations

Units:

°C	Temperature in Celsius
%	Percentage (Number of hundred)
Da	Dalton
kDa	Kilo-Dalton
h	Hour
min	Minutes
s	Seconds
nm	Nanometer
cm	Centimeter
m	Meter
mM	Millimolar
μM	Micromolar
nM	Nanomolar
fM	Femtomolar
M	Molar
μg	Microgram
g	Gramm
kg	Kilogram
l	Liter
ml	Milliliter
ppm	Parts per Million
ppb	Parts per Billion
rpm	rounds per minute
μl	Microliter
V	Volt
mV	Millivolt

Chemical abbreviations:

AA	Acetic acid
ABC	Ammonium bicarbonate
ACN	Acetonitrile
API	Application Programming Interface
ddH ₂ O	Deionized water
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDTA	Ethylendiamintetraessigsäure
EtOH	Ethanol

FA	Formic acid
H ₂ O	Water
HCl	Hydrochloric acid
IAA	Iodoacetamide
MeOH	Methanol
MgCl ₂	Magnesium chloride
NaCl	Sodium chloride
PVDF	Polyvinylidenfluorid
SDS	Sodiumdodecylsulfate
TFA	Trifluoroacetic acid

General abbreviations:

2D	Two dimensional
AD	Alzheimer's Disease (Morbus Alzheimer)
ABI	Acquired brain injury
BSA	Bovine serum albumin
CID	Collision induced defragmentation
CNS	Central nervous system
CRP	C-reactive protein
CT	Computer tomography
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DOI	Digital Object Identifier
DMEM	Dulbecco's Modified Eagle Medium
ER	Endoplasmic Reticulum
ESI	Electrospray ionization
ESR	Erythrocyte sedimentation rate
ETD	Electron Transfer Dissociation
FAB	Fast Atom Bombardment
FASP	Filter-assisted sample preparation
FT-ICR	Fourier transform ion cyclotron resonance
FWHM	Full width at half maximum
γ-GT	Gamma-glutamyl transferase
GO	Gene Ontology
HCD	Higher-energetic Collisional Dissociation
HeLa	"Henrietta Lachs" ovarian cancer cell line
HPLC	High Performance Liquid Chromatography
UHPLC	Ultra-High Performance Liquid Chromatography

HbA _{1c}	Glycated hemoglobin
HDL	High density lipoprotein
iBAQ	Intensity-based absolute quantification
IRB	Institutional Review Board
ICB	Intracranial bleed
IgG	Immunoglobulin G
IHC	Immunohistochemistry
IT	Injection time
LC	Liquid Chromatography
LC-MS	Liquid Chromatography coupled mass spectrometry
LDL	Low density lipoprotein
mAb	monoclonal antibody
Max	Maximal
MRI	Magnetic resonance imaging
MS	Mass Spectrometry
MWCO	Molecular weight cut-off
NCE	Normalized Collision Energy
ND	Neurodegenerative disorders
PCA	Principal Compound Analysis
PD	Parkinson's Disease (Morbus Parkinson)
PDMS	Plasma Desorption Mass Spectrometry
PHP	Hypertext Preprocessor
PMCID	PubMed Central Identifier
PNS	Peripheral nervous system
PP	Polypropylene
PTM	Post-translational modification
PVS	Persistent vegetative state
QE	Thermo Fisher Orbitrap: Q Exactive (mass spectrometer)
R	Programming language for Statistical Analysis
RNA	Ribonucleic acid
ROC	Receiver Operating Characteristic
RT	Room temperature
SDS-Page	Sodiumdodecylsulfate polyacrylamid gel electrophoresis
SNP	Single nucleotide polymorphism
TBI	Traumatic brain injury
TOF	Time-of-flight

TSH	Thyreotropin
UC	Ultra centrifugation
WB	Western Blot
WHO	World Health Organization

(blank page)

1. Introduction

1.1 Mass spectrometry based proteomics

Traditionally, the term “Proteomics” has been associated with the large-scale determination of gene function by biochemical methods¹⁻³. The rise of proteomics can be dated to the late 1970s, when researchers started cataloguing and building databases of two dimensional gel electrophoresis outcomes^{2, 4-7}. However, researchers had to overcome many challenges with these experiments, such as high complexity of cellular proteomes as well as the low abundance of proteins. This in turn presented the necessity for highly sensitive analytical methods^{1, 8}. By the 1990s Mass Spectrometry (MS) had already emerged in the biological research arena, as it overcame many limitations of protein analysis, as e.g. sensitivity^{2, 9, 10}. Mass spectrometry has thus far evolved into an indispensable tool for protein analysis and is clearly the method of choice regarding the analysis of complex protein samples. Due to this recent development in the proteomics technology, the analysis of single proteins (hypothesis driven) via classical biochemistry approaches has evolved into a more holistic, i.e. systematic-driven analysis type, as researchers nowadays have the ability to study the whole proteome at a time instead of a single gene product^{3, 11, 12}. With this highly sensitive technology at hand, the proteomic analysis is not only focused on the identification of gene products or cellular functions, but also on post-translational modifications (PTMs), protein structure, protein function, protein-protein interactions and abundance level changes^{1-3, 8, 12}. Based on the scope of a proteomic study, the applied approach can be classified as one of three different approaches: *top-down*¹³, *middle-down*¹⁴ and *bottom-up*¹⁵.

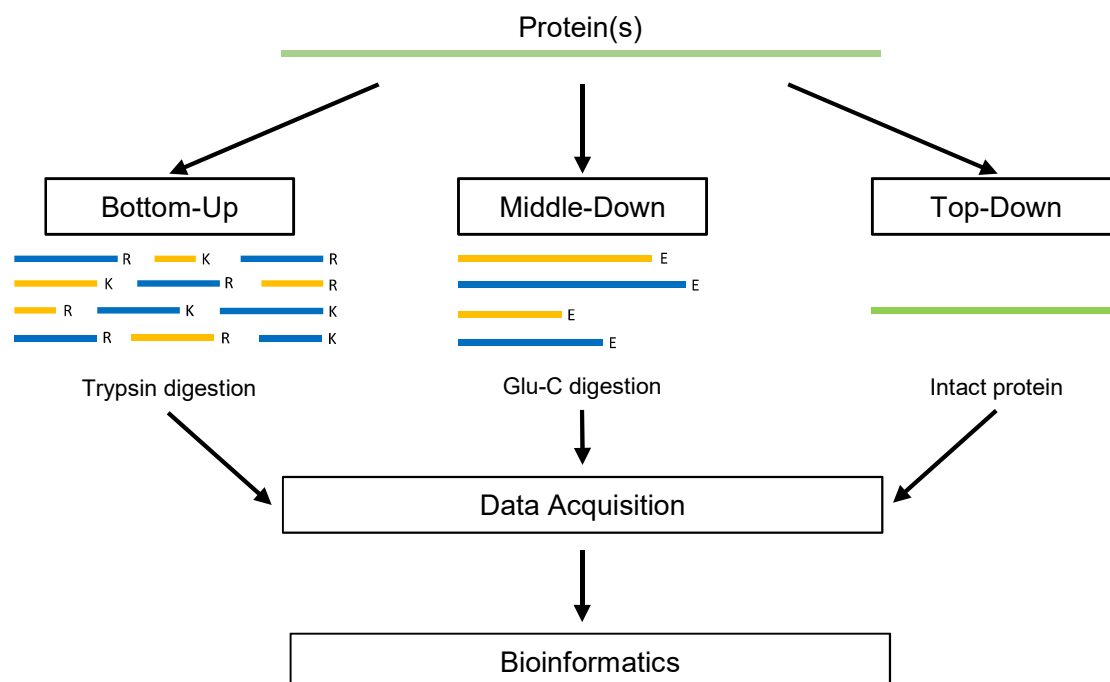


Figure 1.1: Proteomic Workflows. Schematic illustration of the different proteomic workflows. Bottom-Up (left), Middle-Down (middle) and Top-Down (right) represent the most frequently used approaches. Figure was rebuilt in dependence of Moradian *et al.* (2014)¹⁶.

In all three cases, the starting material is a complex protein mixture obtained from a biological sample (e.g. tissue extract) or clinical sample (e.g. body fluid). The main characteristic for the *top-down* approach is the analysis of intact proteins by MS. In this regard, the protein mixture is first separated by chromatographic methods, e.g. 2D - electrophoresis, to select for individual proteins for further MS analysis and protein identification ¹³. In contrast hereto, the *bottom-up* as well as the *middle-down* approach apply proteolytic digestion directly to the whole complex protein mixtures. Consequently, the generated peptide mixture has to be extensively chromatographically separated and analyzed by MS. Here, the length of proteolytic peptides differentiates between the two approaches. While the bottom-up approach mainly relies on protein digestion via e.g. Trypsin, generated peptides feature a size in the range of 0.5 to 3.5 kDa ¹⁷. In direct comparison, the middle-down approach relies on proteases that generate peptides at least twice the size (> 6.5 kDa) of those generated by e.g. Trypsin ¹⁴. However, the bottom-up approach, like the middle-down approach, can't provide full protein sequence information either and hence, suffer from the fact that parent gene products are generally indistinguishable in the absence from additional information, especially since the *de novo* sequencing problem by MS has not yet been fully solved

^{1, 15, 18}.

1.1.1 Mass Spectrometry

By definition, MS measures the mass to charge ratio of an ionized analyte in the gas phase and reports the result in a mass-to-charge (m/z) ratio¹⁹. Additionally, a Mass Spectrometer can be subdivided into an ion source, a mass analyzer and a detector. In many proteomic setups, but particularly those based on the complexity of the protein mixture to be analyzed and the employed ionization technique, the MS system is online coupled with a preceding separating liquid chromatographic (LC) system (LC-MS)^{1, 19}.

1.1.1.1 Liquid chromatography based peptide separation

The proteolytic digestion of a complex protein mixture with the assistance of proteolytic enzymes (e.g. Trypsin) leads to an even higher complex peptide mixture. Theoretically, many peptide species can share the same mass even though they do not share the same progenitor protein they originate from. Consequently, a single acquired mass spectrum of this highly complex peptide mixture will not correctly represent all present peptide species. Hence, peptide separation, based on other chemical or physical properties than its mass, is inevitable. Due to the nature of proteomic samples (small sample volume and low abundant peptide species), liquid chromatography is usually performed on a nano scale

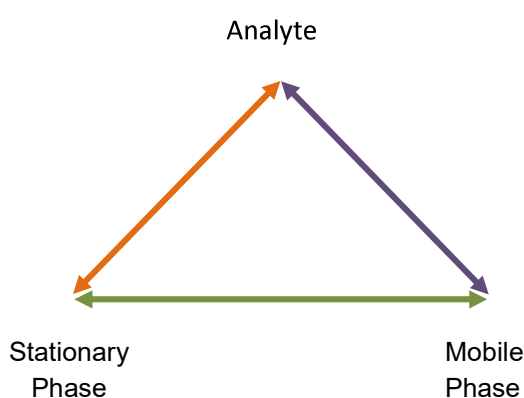


Figure 1.2: Interaction Triangle. The retention time for each analyte is characterized by three interactions: Analyte with Mobile Phase (purple), Stationary Phase with Mobile Phase (green) and Analyte with Stationary Phase (orange).

(ultra-)high performance liquid chromatography system (nano-UHPLC; nano-HPLC)²⁰. This, on the other hand, allows the detection of peptides in the femtomole (fmol) region^{19, 21}. The separation occurs through interaction between the analyte, the stationary phase and the mobile phase (cf. **Figure 1.2**). In most proteomic cases an etherified C-18 silica gel is chosen and hence the separation is based on hydrophobic interactions between the peptide side chains and the stationary phase. By adding ion-pair reagents as trifluoroacetic acid (TFA) or formic acid (FA) hydrophilic side chains can also interact with the embedded amphiphiles in the stationary phase. Based on these characteristics, the frequently employed

peptide species can share the same mass even though they do not share the same progenitor protein they originate from. Consequently, a single acquired mass spectrum of this highly complex peptide mixture will not correctly represent all present peptide species. Hence, peptide separation, based on other chemical or physical properties than its mass, is inevitable. Due to the nature of proteomic samples (small sample volume and low abundant peptide species), liquid chromatography is usually performed on a nano scale

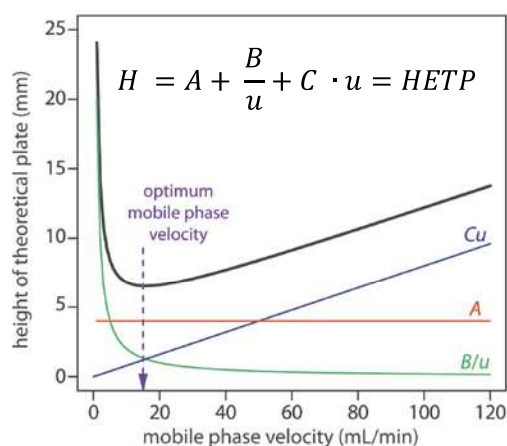


Figure 1.3: van Deemter plot. Combined effect (theoretical plate height, black) for adsorption (grey), Eddy-Diffusion (red) and mass exchange (green). The smaller the theoretical plate height (black) the better the separation effect. Equation: H – plate height (black); A – Eddy-Diffusion (red); B/u – Diffusion (green); C – mass exchange (grey); u – mobile phase velocity (X - Axis); HETP – Height equivalent of theoretical plate. Figure adopted from:

<http://chemwiki.ucdavis.edu>

separation is called “ion pair reversed phase” liquid chromatography (IP-RP-HPLC). In order to elute the analyte off the stationary phase and achieve a separation, the hydrophobic character of the mobile phase is gradually increased by enhancing the content of a hydrophobic solvent, as e.g. acetonitrile (ACN). Once the hydrophobic character reaches a certain level, the analyte interacts preferably with the mobile phase and is no longer retained by the stationary phase. Nevertheless, to achieve an optimal peptide separation and hence a small plate height, the velocity of the mobile phase also has to be tuned for each setting, based on van Deemter *et al.* (1995)²². In this context, the estimation of the flowrate of the mobile phase (*cf.* **Figure 1.3**) not only considers the velocity itself, but also the diffusion of the analyte along the column material in consideration of the package of the column (Eddy-Diffusion), the regular diffusion due to concentration difference along the column (Diffusion) and the actual interdependency between analyte and stationary phase (Adsorption/Desorption).

1.1.1.2 Ionization Techniques

The development of two soft ionization techniques, namely electrospray ionization (ESI) by Fenn *et al.* in 1989²³ and matrix-assisted laser desorption ionization (MALDI) by Karas *et al.* in 1988²⁴, revolutionized mass spectrometric protein and peptide analysis. These two techniques have demonstrated the exemplary ability of ionizing peptides or proteins¹⁹ and remain the most commonly used ionization techniques in the field of proteomics to this day.

Besides ESI and MALDI, further soft ionization methods as “Fast Atom Bombardment” (FAB)²⁵,²⁶ and “Plasma Desorption” (PDMS)²⁷ are available to scientists for protein analysis. These ion techniques are classified as particle bombardment (PB) and are closely related to “Secondary Ion Mass Spectrometry” (SIMS). Due to the high complexity of the aforementioned ionization techniques, only ESI and MALDI are further explained in the following:

Electrospray Ionization

ESI is the ionization method of choice, once the LC is online-coupled to the MS (LC-MS). Additionally, this technique produces ions with multiple charges. For a successful electro spray, a high voltage (> 1.8 kV) is applied between the end of the fused silica capillary (emitter tip) post the separation column and the ion inlet of the MS. The voltage and hence the potential difference causes the formation of a “Taylor Cone” (*cf.* **Figure 1.4**) from which charged parent droplets are emitted. Due to extensive heat or a uniform gas flow (zero-air) applied along the capillary, residual solvent evaporates from the parent droplet. Consequently, the super-charged droplet undergoes “Coulomb fission” leading to smaller highly charged droplets, containing single analyte molecules. The consequent process of analyte charging is further described by two hypotheses²⁸. On the one hand, the passive process of “Charged Residue Model” postulates that the small charged progenitor droplet has further losses of solvent through evaporation until the analyte remains naked. Consequently, the charge is consigned to the analyte. On the other hand, the active process of “Ion Evaporation Model” implies that the analyte is ionized by direct evaporation out of the droplet and hence, drags charges off the droplets surface which in turn ionizes the analyte itself²⁸.

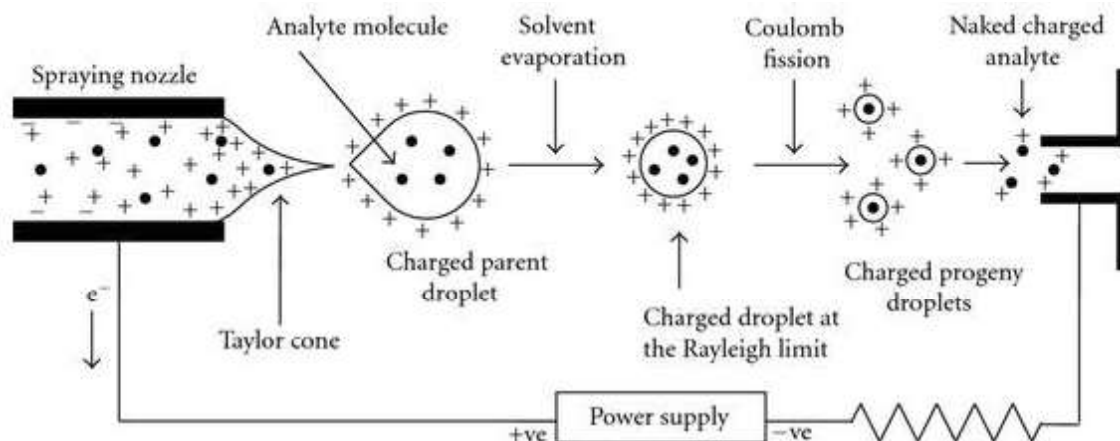


Figure 1.4: Schematic illustration of electrospray ionization. The application of a high voltage (> 1.8 kV) between the emitting end of the fused silica capillary (emitter tip, left) and the ion source of the MS (right) leads to the formation of a “Taylor Cone”. Due to extensive heat or gas flow, the solvent evaporates. The supercharged solvent bubble hence undergoes “Coulomb fission” so that the charge gets transferred to the analyte and a naked charged analyte enters the MS. Figure was adopted from Banerjee *et al.* (2012) ²⁹.

Matrix-Assisted Laser Desorption Ionization

The MALDI ionization technique was developed and introduced by Karas & Hillenkamp in 1988 ²⁴. One of the main characteristic is the applicability to larger biomolecules, i.e. from larger peptides to intact proteins. However, this soft ionization technique is not directly applicable for online coupling to a MS. Here, the sample first has to be prepared on a proper sample plate before it can be used for MS analysis ^{30, 31}. In this case, the desired sample (low complex) is mixed with a 1,000 to 10,000-fold molar excess of a laser wavelength absorbing matrix (small organic molecules, *cf.* **Table 1.1**) directly on the sample plate. After evaporation of the solvent, the analyte co-crystallizes with the matrix and is fully embedded in the matrix crystal structure ²⁸. For ionization, the matrix-analyte crystal is consequently exposed to laser impulses for few nanoseconds. In the case of the laser energy being set too high, the larger analytes (e.g. proteins) will be destroyed immediately. Appropriate lasers are e.g., Nitrogen, Yttrium-Aluminum Garnet, Excited Dimer (Excimer) or Carbondioxide (CO_2). When exposed to radiation, the matrix molecules adsorb the UV radiation which is consequently released into the crystal lattice and causes a lattice disruption and expansion. Hence, the matrix surface gets disrupted and matrix molecules, such as analyte molecules translate into the gas phase. Ultimately, the analyte is ionized due to photo-ionized and radical matrix molecules and hence by their induced proton transfer ²⁸.

Table 1.1: Exemplary list of appropriate matrices for MALDI ionization. Matrices serve the function of laser energy desorption, protecting the analyte from thermal decomposition and ionizing the embedded analyte by proton transfer in the gas phase ^{11, 28}.

Matrix Molecule	Type of Analyte
Sinapinic acid	Proteins
Nicotinic acid	Proteins
Ferulic acid	Proteins

2-(4-Hydroxyphenylazo)benzoic acid (HABA)	Proteins, Peptides
2,3-Dihydroxybenzoic acid (DHB)	Proteins, Peptides
Succinic acid	Proteins, Peptides
α -Cyano-4-hydroxycinnamic acid	Peptides
3-Hydroxy picolinic acid	Oligonucleotides

1.1.1.3 Mass Analyzer

The mass analyzer represents the central part of MS and particularly in the field of proteomics, the mass analyzer requires the following key-features: sensitivity (defines the ability to detect an ion at a certain abundance), mass resolution (defines the ability to distinguish two peaks of slightly different m/z), mass accuracy (defines the error between the measured and true ion mass) and the ability to generate fragmentation spectra of precursor ions (tandem MS or MS/MS; cf. **Section 1.1.1.4**). There are four mass analyzers commonly used in MS-based proteomics, namely i) ion traps, ii) quadrupoles, iii) time-of-flight (TOF) and iii) Fourier transform ion cyclotron resonance (FT-ICR). Each of these mass analyzers can be coupled with different ion sources. Commonly used combinations are e.g. MALDI-TOF, ESI-TOF or ESI-FTICR. Each analyzer varies in design and performance, however each features its own set of strengths and weaknesses. Each analyzer can be used on a stand-alone basis, or in a coupled arrangement so that the strengths of each can be taken advantage of ^{1, 2, 32, 33}. Due to the complexity of the underlying facts, only two of the aforementioned mass analyzers (TOF & Orbitrap) will be introduced briefly.

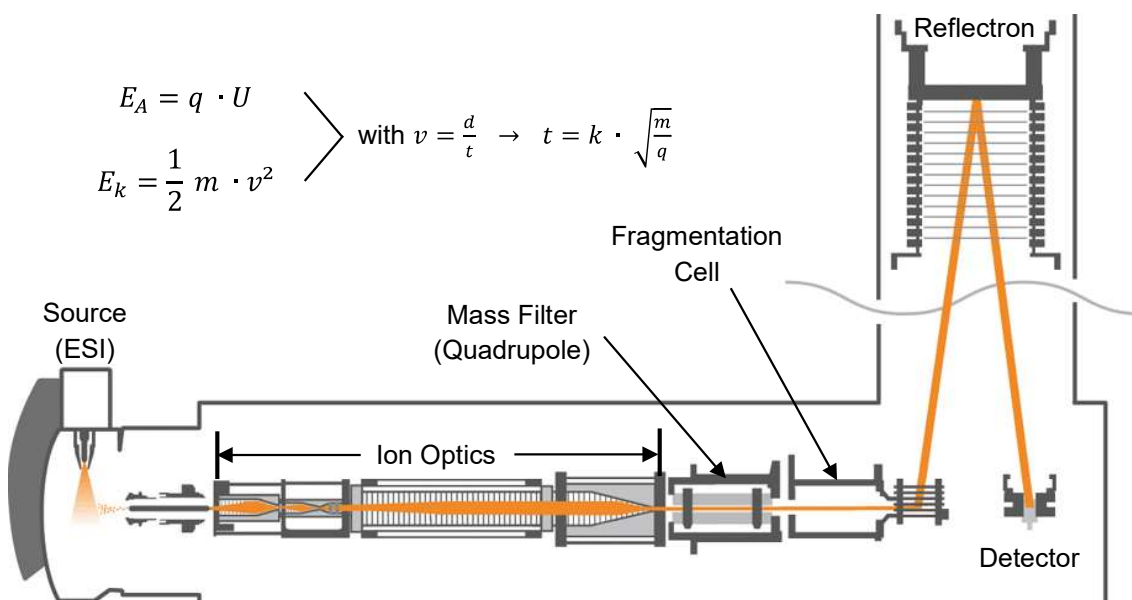


Figure 1.5 Schematic illustration of a time-of-flight mass analyzer. After ESI ionization, the ion beam is condensed through the ion optics. After passing the Quadrupole and fragmentation cell, the ions are quickly captured in a Faraday-cup and then pulsed into the TOF tube. After reflection, the time of flight in the tube is measured by impact on the detector. E_A = Energy of analyte; E_k = Kinetic energy; q = Charge; U = Voltage; v = Velocity; m = Mass of analyte; d = Distance; t = Time; k = Constant of proportionality. Figure taken from <http://www.agilent.com>.

Time-of-Flight (TOF)

As already defined by its name, the “time-of-flight” mass analyzer determines the mass-to-charge ratio of a charged ion by the time needed to drift between a set start and end point (detector). In short, after ionization the charged analytes undergo a kinetic acceleration by traversing an electrical potential. The adsorbed energy equals the product of applied voltage (U) and the charge (z) of the ion ($E = U \times z$). Hence, based on the relation of energy and velocity (cf. **Figure 1.5**) the time required to drift along a predefined path within the analyzer is proportional to the square root of the mass-to-charge ratio. Besides other types, the TOF is not a mass filter, which in turn means that all ions are measured simultaneously. Hence, the necessity for discontinued mass range scanning is omitted which allows a very fast mass spectra acquisition (theoretically up to thousands per second). Cutting edge TOF instruments can achieve measurements with an accuracy error of 2 ppm and a resolution of up to 30,000. Nevertheless, such high mass resolution can be achieved by specifically extending the flight path of the ions. Here, the elongation is technically realized by slowing the ions down at the end of the first leg by applying a voltage between parallel-arranged metal plates, which in turn is inverted in comparison to the actual flight path. Consequently, entering ions into either a linear or staged electrical field are first slowed down and then re-accelerated in the opposite direction. Ions behave like “light” in this context and hence can undergo specific flight path manipulations, so that the ions are specifically directed towards the detector. Here, the specific arrangement of metal plates is also called “Reflectron”, especially based on the reflecting potential as ion mirror^{28, 34-36}.

Orbitrap Mass Analyzer

In the year 2000, Alexander Makarov introduced the new technology of the Orbitrap analyzer³⁷. This technology was consequently adopted by Thermo Fisher Scientific (<http://www.thermofisher.com>) and made commercially available to the community five years later as part of a hybrid instrument³⁸. Different types of Orbitrap analyzers have thus far evolved and are built into different models or hybrids (e.g. Orbitrap Fusion, QExactive, Velos, Elite, etc.) based on their strengths.

The Orbitrap mass analyzer consists of an inner and outer electrode, while the inner electrode features a spindle-like shape (cf. **Figure 1.6 (a)**). For the measurement of the m/z ratio, an ion package consisting of different mass species is injected into the electric field (between inner and outer electrode) of the Orbitrap from an upstream-embedded so-called “c-trap”. Right after the ions enter the Orbitrap’s electric field, the ions start moving in three different directions:

i) oscillation between inner and outer electrode, ii) oscillation along the inner electrode and iii) rotation around the inner electrode (φ). After reaching a “steady state”, different ion species oscillate as an orbit along the z-axis of the inner electrode. Finally, the different ion rings are detected by the image current caused in the outer electrode by passing rings. The characteristic spectrum is achieved by deconvolution of the acquired image current by Fourier transformation. A higher mass resolution can easily be achieved by either strengthening the electric field within the Orbitrap or by prolonging the acquisition time. The Orbitrap data acquisition is directly comparable to the FTICR technique, while the difference is solely based on the absence of a magnetic field^{37, 39-41}.

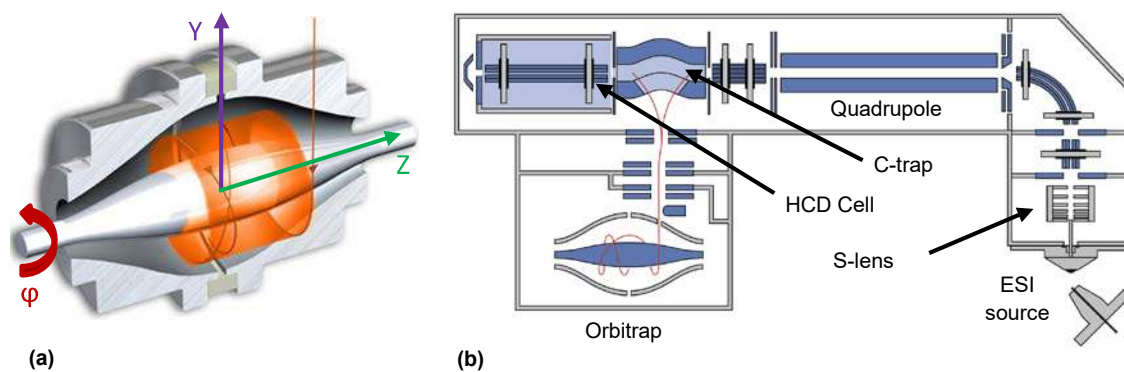


Figure 1.6: The Orbitrap mass analyzer. The m/z ratio is measured by acquiring the image current of the frequency of ion packages oscillating. **(a)** Schematic illustration of an Orbitrap mass analyzer. After injection of ions from the c-trap into the Orbitrap (orange), ion form orbitals that move in three different directions (Y, Z and ϕ). **(b)** Schematic built-up of a Thermo Fisher Scientific QExactive Orbitrap Instrument. After ionization by an ESI source, a condensed ion beam is formed by leading ions through the S-lens and a bended quadrupole. The embedded quadrupole functions as mass filter for specific precursor fragmentation in the HCD fragmentation cell. The c-trap functions either as a transfer electrode or as injector to the Orbitrap.

1.1.1.4 Tandem Mass Spectrometry

The acquisition of peptide mass fingerprints might allow for the identification of their originating protein. However, no sequence-specific information can be gathered. Here, peptides can share the same composition of amino acids and are therefore equiponderate, but the difference in their amino acid order let them originate from completely different parent proteins. Hence, protein identification from a complex or even highly complex protein mixture, simply based on the acquired peptide mass fingerprint will turn out quite inaccurately. Additionally, in regard of PTM analysis, knowing from which amino acid the modification originates or which sequence tag has been modified is indispensable in tracing back to biological functions^{1, 8, 18, 19}.

The principal idea of tandem MS, i.e. MS/MS is the acquisition of a mass spectra of a fragmented precursor ion. Therefore, a regular mass spectra of a present peptide species is acquired (MS^1). Based on the ion intensities in the acquired MS^1 , a preset number of intense ions is selected for fragmentation (e.g. TOP 10, the 10 most intense ions are selected for successive fragmentation). Consequently, the embedded mass filter (“quadrupole”) selects for only one specific mass species at a time. After passing the mass filter, the ion package is transferred into a “Collision cell”. After fragmentation, the generated ion fragments enter the mass analyzer and a second mass spectra is acquired (MS^2). Based on the nature of the ion fragments, the mass difference between monoisotopic mass peaks allows for the determination of amino acid residues. By following the mass difference and hence the progression of amino acid residues from light to heavy ions, the peptide sequence can be determined^{1, 28, 33, 42}.

There are various different ways of ion fragmentation in the gas phase, but for proteomic samples, *collision induced dissociation* (CID)⁴³, *higher-energy collisional dissociation* (HCD)⁴⁴ and *electron transfer dissociation* (ETD)^{45, 46} are most commonly used.

Using the CID method, preselected ions are further accelerated by an electric field after passing the mass filter. Consequently, ions enter a fragmentation cell (quadrupole or octopole) filled with neutral inert gas (Nitrogen, Helium or Neon). Within this cell, the analyte ions collides with the gas

molecules. Upon collision, the kinetic energy of the analyte ion is converted into vibrational energy, resulting in backbone breakage and ion fragmentation ⁴⁷.

The HCD fragmentation method in principal, is identical to the CID method. Nevertheless, the HCD method is only used in orbitrap mass spectrometers. Hence, the HCD method is also synonymously called *higher-energy C-trap dissociation*. Here, the analyte ions are transferred through the C-trap (cf. **Figure 1.6**) into the collision cell. After fragmentation, the fragments are again transferred back to the C-trap for injection into the orbitrap. Here, due to the higher energies used, a higher fragmentation rate can be achieved as well as the detection of ions in a lower m/z range compared to CID ⁴⁴.

Besides collision-induced fragmentation (CID/HCD) peptide ions can also be fragmented by transfer of an electron. Hereby, analyte ions are transferred into a fragmentation cell (quadrupole/octopole) and a chemical ionized reactive anion radical is consequently introduced (anthracene [C₁₄H₁₀]; fluoroanthene [C₁₆H₁₀]) into the fragmentation cell. Finally, after the transfer of an electron to a positively charged peptide ion causes breakage along the backbone ^{46, 48}.

Based on the nomenclature for peptide fragment ions (cf. **Figure 1.7**) ⁴⁹ CID and HCD fragmentation methods preferably generate b_n and y_n ions. Otherwise, the ETD fragmentation method preferably produces c_n and z_n ions ⁵⁰. For de-novo sequencing of peptides by MS² spectra an assumption of present ion types is essential.

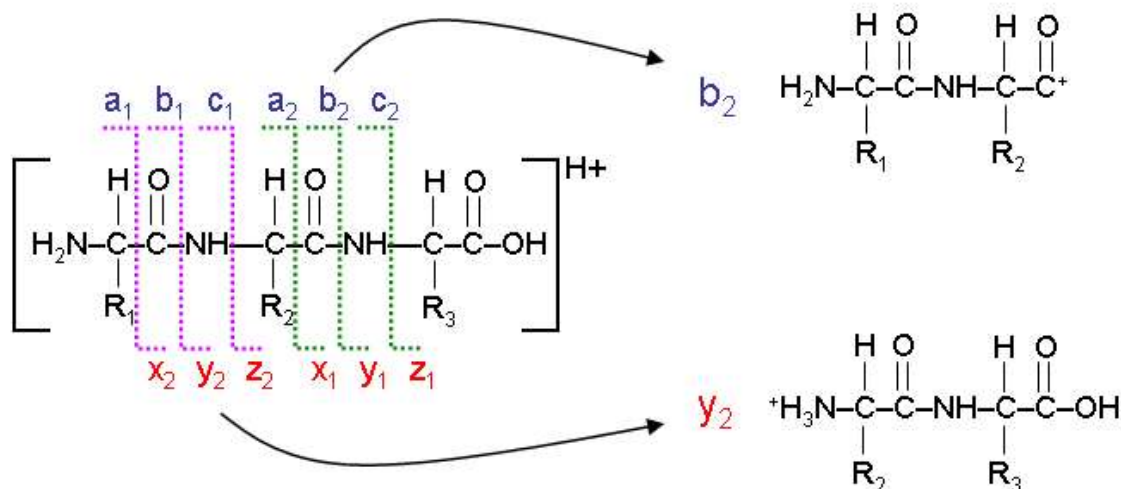


Figure 1.7: MS-based peptide backbone fragmentation. Fragmentation causes a peptide backbone breakage. Based on which fragmentation technique was used, different fragments are generated according to preference. The ions are consequently defined by the position on which the charge resides after fragmentation. An amino acid sequence can be easily determined by b_n and y_n ion series. Here, the charge resides either on the N- or C-term of a peptide fragment. The figure was adopted from <http://www.alchemistmatt.com>.

1.1.1.5 Protein Identification

An important step for protein mass spectrometry is the identification of proteins out of the acquired data. There are two different ways for protein identification: i) Peptide Mass Fingerprinting or ii) De Novo Peptide Sequencing.

Peptide Mass Fingerprinting

Here, the acquired peptide masses (MS^1) function as input for this identification approach. Consequently, the acquired MS^1 masses of e.g. tryptic peptides are compared to *in silico* generated tryptic peptides originating from a protein sequence database. A protein can be considered “identified” once a significant number of experimental masses match the computed reference masses. Nevertheless, for better and accurate matching of experimental and reference masses, protein purification represents an inevitable step and limits its use for higher throughput approaches. Peptide Mass Fingerprinting is hence widely used for protein identification with preceding digestion within top-down proteomic approaches ⁵¹.

De Novo Peptide Sequencing

The acquisition of MS^2 spectra of peptide species allows for the determination of the underlying peptide sequence (*cf.* **Section 1.1.1.4**). Hereby, the occurring mass difference between peptide fragments in the MS^2 spectra is used, to directly associate a specific amino acid at the given sequence position. Given the consecutive mass differences, the peptide sequence can be deviated. Based on the found sequence, the identified peptide can consequently be assigned to either a parent protein or protein group. In some cases, the determined sequence is shared amongst multiple proteins. The corresponding peptide is thus considered a “shared peptide”. In turn, peptides that demonstrate a sequence which is unique for only a single protein is thus called “unique peptide” ^{1, 33}. In theory, a protein can be considered “identified” once a corresponding unique peptide has been detected. Nevertheless, a “two peptide” rule is widely applied ⁵².

An accompanying challenge with the search of mass spectra or tandem mass spectra against a reference database, imposes the identification of “false-positive” hits. This effect can be specifically avoided, i.e. addressed by an estimation of a “false discovery rate” (FDR) for each acquired dataset ⁵²⁻⁵⁴. This in turn, allows for filtering and minimizing the occurrence of such false positive hits. The FDR is estimated by searching the acquired mass spectra against a database containing original target protein sequences as well as decoy sequences (e.g. scrambled or reversed protein sequences) ^{55, 56}.

1.1.2 Quantitative Mass Spectrometry

Besides only providing a list of identified proteins, mass spectrometric data can also provide more specific information about the abundance of identified features. In this regard, the signal intensity, which a certain peptide ion or fragment ion causes, gives rise to its biological abundance in the actual biological system the sample originates from⁵⁷⁻⁵⁹. There are two approaches which are generally used to quantify proteins or peptides by MS: label-free and label-based quantification.

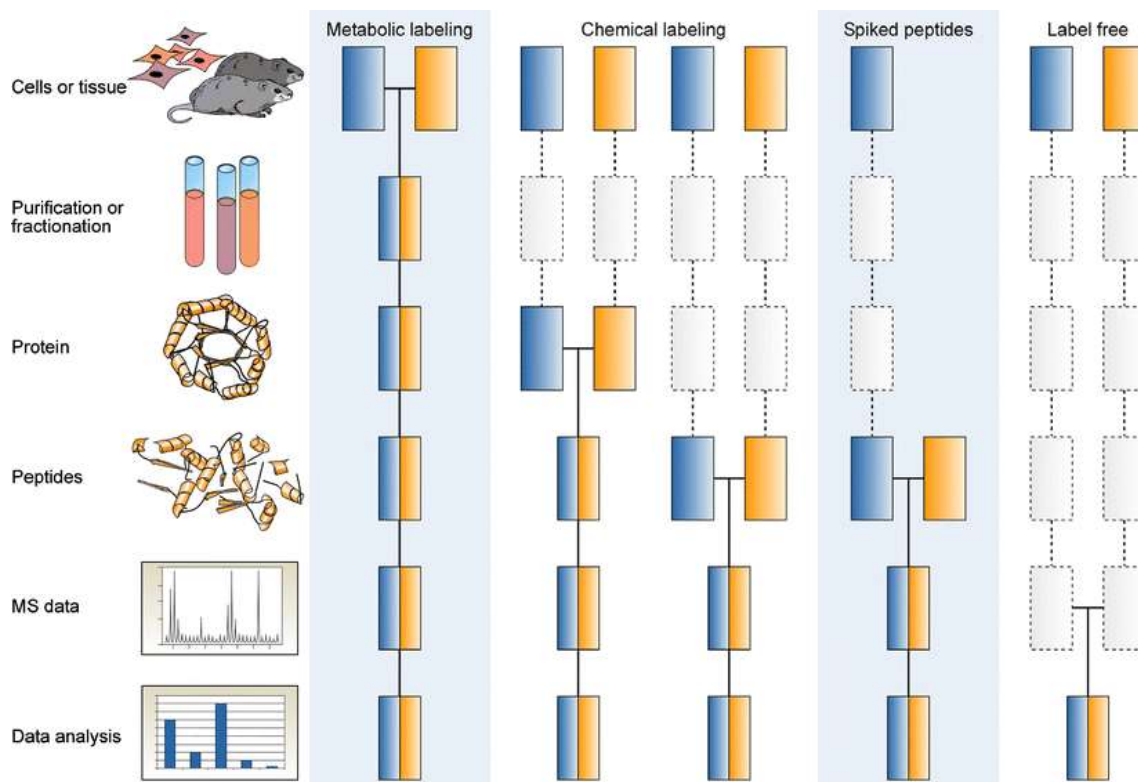


Figure 1.8: Quantitative proteomic workflows. MS quantification occurs on a relative basis. The earlier two different conditions can be combined for analysis, the better the comparability between LC-MS runs. Label-based quantification workflows (Metabolic labeling, Chemical labeling and Spiked peptides) allow for combination either already on the sample, protein or peptide level. Despite this, label-free approaches only allow for combination on the MS data level. Figure adapted from Bantscheff *et al.* (2007)⁵⁷.

1.1.2.1 Label-based Quantification

Ideally, each sample is processed identically to be comparable to each other. Hence, the best way to technically realize this scenario is to combine samples at the earliest possible stage in the workflow, so that only one sample has to be either processed and/or analyzed which in turn leads to the fact that an introduced error affects each individual sample equally.

In this context, labeling proteins or peptides allows for the combination of samples at their earliest intercept but especially with the ability to retrieve quantitative information individually for each pooled sample on the MS-data level.

The label-based quantitative mass spectrometry approach makes use of a specifically introduced mass drift of identical peptide or protein species by the incorporation of isotopically different amino acids. Here, isotopes do change the actual mass of a protein or peptide but does not influence

any chemical properties associated with it. Hence, isotopically labeled proteins and peptides can be seen in the same mass spectra simultaneously with either its normal or other labeled specie. However, the actual mass spectra interpretability will suffer from a higher complexity as identical protein or peptide species are represented by multiple peak clusters. Nevertheless, based on the aforementioned mass shift, the peak clusters can be mapped exactly back to each sample which was pooled prior to analysis. A distinction is generally made between three different methods for labeling protein/peptides: (i) metabolic (protein level) (ii) chemically (protein or peptide level) (iii) spike-in (peptide level).

Metabolic labeling

The metabolic labeling approach describes the incorporation of isotopes into proteins during cell growth/division *in vitro* (cell culture) but also *in vivo* (animal model). Human samples, or clinical samples in particular, cannot be metabolically labeled due to ethical discrepancies, unless samples can be further cultivated *in vitro* and labeled there.

Incorporation of isotopes into the whole proteome is realized by either culturing cells in a medium with isotopic nutrition (e.g. ^{15}N) for multiple cell cycles or by already providing “stable isotope labeled amino acids” (SILAC) in the medium. By providing isotopically labeled substrates already on the metabolic level, the incorporation of such isotopes theoretically occurs with the whole proteome and each protein should appear with the specific mass drift. After MS analysis, proteins and peptides can be quantitatively compared by relatively linking the MS¹ spectra signal intensities.

Chemical labeling

Besides incorporation of isotopes into the whole proteome by metabolic processes, proteins and peptides can also be chemically or enzymatically labeled, even after extraction out of its original environment. Hence, this type of protein/peptide labeling approach is also considered as post-biosynthesis labeling.

An elegant way to introduce isotopes into proteomic samples via an enzymatic reaction is by exploiting the catalysis of the oxygen exchange reaction by class-2 proteases (e.g. trypsin)⁶⁰⁻⁶². The incubation of tryptic peptides with immobilized trypsin and water with an isotopic oxygen, i.e. H_2^{16}O leads to an exchange of two oxygen atoms at the C-terminal carboxyl group and in turn leads to a mass shift of 4 Da⁶², but more importantly to no changes of any chemical preference. Apart from enzymatically labeling peptides, there are multiple options to chemically label proteins or peptides for mass spectrometric quantification.

Firstly, Gygi *et al.* proposed a method in 1999⁶³ to specifically derivatize cysteine residues with an agent containing either no or up to eight deuterium ions. Nevertheless, after an initial success, it turned out that the method might not be as suitable for peptide/protein quantification as thought, as cysteine residues might either be missing or only be represented by a few residues^{57, 63}.

Secondly, proteins and peptides are commonly quantified by using isobaric labelling tags, namely either “Tandem Mass Tags” (TMT) or “isobaric tags for relative and absolute quantification” (iTRAQ). This method uses *N*-hydroxysuccinimid (NHS) chemistry to specifically derivatize the

N-terminus of a protein or peptide but also both lysine and tyrosine (only iTRAQ) side chain residues. These tags do not cause a mass shift for each sample specie they are covalently bound. Hence, the MS¹ spectra does not suffer from any further complexity. During the fragmentation step in the mass spectrometer, the "reporter group" of the attached and mass balanced tag is split away and can be acquired as specific ion in the corresponding reporter region in the tandem mass spectra (MS²): 125 to 132 m/z for TMT (6-plex), 126 to 131 m/z for 10-plex TMT and 112 to 122 m/z for iTRAQ (8-plex). For quantitative analysis between the pooled samples, the acquired ion intensities of the corresponding reporter ions (1 Da mass shift between each reporter ion) can be compared relatively. While the TMT tag (6 reporter ions at 126, 127, 128, 129, 130 and 131 m/z) allows for up to 6 concurrent samples being pooled, the use of iTRAQ (8 reporter ions at 113, 114, 115, 116, 117, 118, 119 and 121 m/z) tags allows for up to 8 simultaneous samples.

Lastly, the N-terminus as well as the lysine side chain residues of proteins or peptides can also be derivatized with dimethyl groups in diverse mass configurations. This technique allows for pooling three different samples/conditions (light [+28 Da], intermediate [+32 Da] and heavy [+34 Da]) by exchanging both hydrogen residues with deuterium and the carbon residue with heavy carbon (¹³C). Signal intensities are consequently extracted from the three different peak clusters in the acquired MS¹ spectrum.

Spiked peptides

Quantification via the aforementioned techniques occurs on a relative basis. Hence, the result reflects only the relation between the pooled and simultaneously analyzed samples/conditions. However, in some cases, the absolute concentration of an analyte has to be described. An absolute quantity estimation is technically realized by spiking an internal standard into the MS-ready sample (peptide level) in a known concentration. Here, either stable isotope-labeled standard peptides, i.e. natural occurring peptides with heavy amino acids or in some cases even artificial peptides are used. Artificial peptides have the advantage over labeled natural peptides in that they can be specifically detectable and also can't interfere with their natural occurring counterpart. After all, the comparison of the signal intensity between target feature and standard peptide allows the absolute concentration estimation and hence for absolute quantification (AQUA). Nevertheless, comparing either spiked peptide intensities for AQUA between separate MS analyses or calculated absolute quantities between runs, remain a critical endeavor.

1.1.2.2 Label-free Quantification

Apart from labeled and pooled samples, non-labeled and individual MS analyses can also be used for quantification. Here, the relative comparison between MS runs can only be performed on the MS-data level (*cf.* **Figure 1.8**). This also imposes the latest point of possible combination. Due to the fact that samples are processed and analyzed separately from each other, many inter-run variances can be introduced simply by different sample handling or altered LC conditions due to column impurities from previous runs. These technical changes can in turn cause statistical significant changes within the acquired signal intensities. Therefore, stringent intensity normali-

zation has to occur before a relative quantitative comparison between label-free MS-runs becomes feasible. At present, there are three widely used label-free quantification strategies: (i) spectral counting, (ii) intensity-based label-free quantification and (iii) data-independent acquisition.

Spectral counting

As mentioned before, the regular data acquisition (data-dependent acquisition; DDA) acquires MS¹ spectra out of which a specific number of precursor ions are selected for consecutive fragmentation (CID/HCD) and MS² spectra acquisition. The MS² spectra, which are acquired, are searched accordingly against a database containing target and/or decoy sequences via e.g. MaxQuant, Mascot, ProteinPilot, MSGF+, etc. This in turn allows for peptide sequence identification for every such MS² spectrum submitted and consequently leads to protein or protein group assignment. Each spectrum that can be uniquely assigned to a protein is hence “counted” amongst its abundance. Spectra that represent a shared sequence should however be excluded for consideration. Consequently, spectral counting can only be used for protein quantification whereas the quantitative value is represented by an integer number of assigned MS² spectra with an identified unique peptide sequence of the parent protein. Nevertheless, the acquisition of MS² spectra is based on a stochastic event, which might bias the quantitative outcome especially at low abundant proteins as peptide ions might not be selected for fragmentation and MS² acquisition by the mass spectrometer algorithm because of constant ion ranking below the set intensity threshold.

Intensity-based label-free quantification

In contrast to spectral counting, the intensity-based label-free quantification approach makes use of the acquired signal intensities in the MS¹ spectrum. After identification of a peptide sequence through MS² spectrum analysis, the signal intensity is integrated along the extracted ion chromatogram (XIC) of the according precursor mass specie. In short, the signal intensity is summed across the corresponding peptide elution peak. Using this approach and hence considering the signal intensity of the corresponding peptide specie not only proteins, but also single peptide species can be easily quantified.

Particularly for protein quantification, either only the signal intensities for each unique peptide can be summed or the “intensity based absolute quantification” (iBAQ) approach can be applied. While the regular label-free intensity based approach considers only the sum of peptide signal intensities across their elution peak, the iBAQ strategy also normalizes the number of peptides used for quantification (i.e. only non-shared peptides).

Data-independent acquisition

The aforementioned label-free quantification strategies are performed on the DDA data level already acquired. The data-independent acquisition (DIA) strategy however, describes a novel acquisition routine, which is programmed for the fragmentation of all precursor ions and the acqui-

sition of a consecutive MS² mass spectrum. In short, the whole desired mass range is consecutively covered with extraction windows while in some cases the window size is set to a fixed size (e.g. 25 Da) in other cases the windows size can be variably set along the mass range. Consequently, all precursor ions detected in the MS¹ and those which fall into one extraction window will be simultaneously fragmented and one MS² spectrum is acquired which then represents one particular extraction window at a given retention time.

Nevertheless, the actual data analysis and hence peptide or protein quantification requires previously acquired DDA data from the same sample type, to be able to correctly identify each fragmented precursor. Particularly with complex samples, multiple precursors are fragmented within one extraction window and their fragment ions are then acquired in a single MS² spectra. Given this situation, a simple database search of this MS² spectrum to correctly identify the involved peptide precursor remains nearly impossible. Only with extensive extrapolations and specialized software (e.g. DIAUmpire) approximated and simplified MS² spectra can be searched against a protein sequence database.

To avoid extensive extrapolations and eventual deconvolutions, the identification results originating from data acquired in the DDA style are stored in so called "ion libraries". These libraries allow the analyzing software to re-identify the fragmented precursor species in the DIA MS² spectrum based on mapping retention time and peak patterns with identified features and hence similar preferences.

This quantification approach calculates the sum of MS² signal intensities across acquisition cycles for quantification. The MS² intensities are considered more stable against technical interferences and hence can be compared better across multiple LC-MS analyses when using this described acquisition technique.

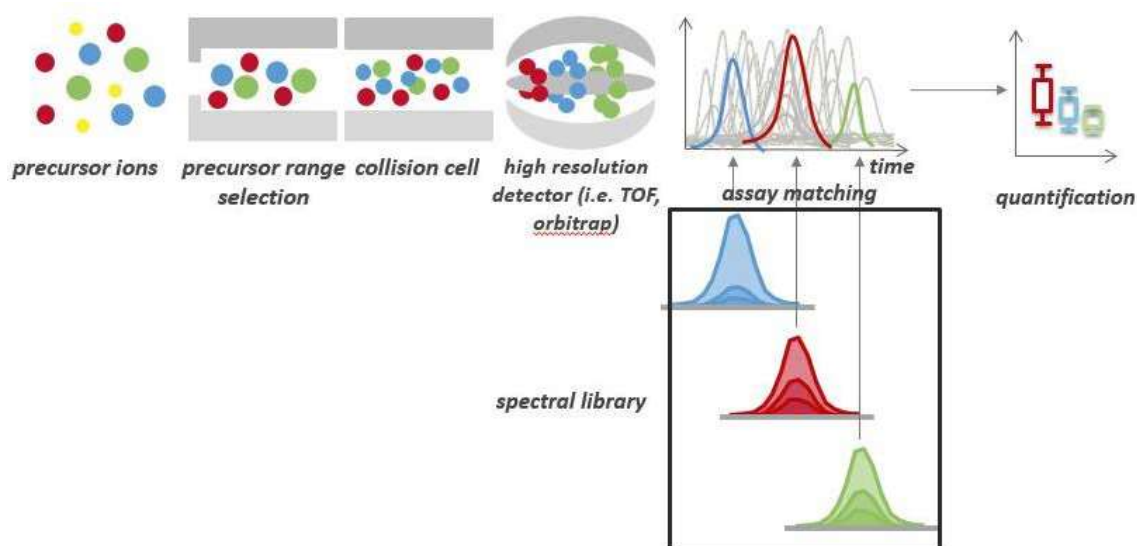


Figure 1.9: Data-Independent Acquisition. Instead of selecting precursors for fragmentation by their respective abundance, while running within data-independent mode, the completely defined mass range is consecutively selected for fragmentation within e.g. 25 Da windows. Consequently, the acquired MS² spectra might be of fragments originating from multiple precursor ions. Previously acquired and identified MS² spectra (i.e. data-dependent acquired MS² spectra) can then be matched to specifically identify involved precursor ions. Figure taken from <http://www.researchgate.net>.

1.2 Biomarker

By definition, a biomarker is “a naturally occurring characteristic (e.g. metabolites, gene, gene transcript, proteins, etc.) or a molecule which is obviously measurable and evaluated as an indicator of normal biological, pathogenic or pharmacological processes”. Biomarkers are especially used in clinical settings to objectively assess the health of human beings. Here, blood pressure or the cholesterol level in the blood of a patient is used for objective assessment. Nevertheless, the use of biomarkers has already been known for centuries. Since antiquity, doctors or healers used occurring discrepancies of certain characteristics (e.g. mucus consistency) for diagnosing diseases. During this time, healers actually investigated a patient’s urine for changes in taste, smell or color. Any deviation from their known “healthy” state led to the association of an irregular incident as well as to treatment with methods that were known then. To this day, urine and its analysis still represent a valuable source for diagnosing diseases or biological conditions. One of the best known examples is the test for presence of hCG (human Chorion Gonadotropin) in female urine and the consequent diagnosis of a possible consistent pregnancy. Additionally, the analysis for the presence of biologically secreted glucose in the urine and its correlation to the assumption that the patient suffers from diabetes mellitus, imposes another important example for the use of biomarkers in a clinical setting.

Table 1.2: Exemplary biomarkers used in the clinical setting. Abbreviations used: HbA_{1c} - Glycated hemoglobin; IgG – Immunoglobulin G; γ -GT - Gamma-glutamyl transferase; LDL – low density lipoprotein; HDL – high density lipoprotein; T3/T4 – Thyroid hormones; TSH – Thyreotropin. Table after Bracht (2009).

Disease	Clinically applied biomarkers
Diabetes mellitus (DM)	Glucose, Ketone, HbA _{1c} , Insulin
Multiple Sclerosis (ED)	IgG, Myelin
AIDS (HIV)	CD4 cells, Virus hull proteins, RNA particles
Kidney dysfunction	Creatinine, Urea, Electrolytes (Na ⁺ , K ⁺ , Cl ⁻)
Liver dysfunction	Bilirubin, γ -GT, LDL / HDL ratio, Hepatitis virus particles
Thyroid dysfunction	T3/T4, TSH
Osteoporosis	Calcium, Phosphate, Alkaline phosphatase
Rheumatoid Arthritis	Rheumatism factor, ESR, C-reactive protein (CRP)

Nevertheless, biomarkers are not only coupled with a binary event (e.g. present vs. absent) but also to general concentration/abundance variations in particular, and thus based on statistical significant quantitative changes. Quantitative changes (higher-than-normal or lower-than-normal) of such biomarker substances are generally investigated in either body fluids (e.g. blood, urine, CSF, etc.) or tissue biopsies, especially in medical settings. In this context, biomarker molecules are represented by single nucleotide polymorphisms (SNPs), DNA methylation patterns, RNA or mRNA changes, but proteins or metabolites also demonstrate abundance variations. Ideally, body

fluids that contain the desired biomarker candidate should be easily accessible via non-invasive (e.g. urine, saliva or tears). Nevertheless, some biomarkers require semi-invasive procedures (blood, CSF, amniotic fluid) or even tissue biopsies to be accessible for analysis. The clinical applicability of biomarkers requires high sensitivity (detectability of changes also at very low biological concentrations) as well as specificity (observed changes are truly related to the diseased state).

1.2.1 Types of biomarker

Based on the medical background or clinical application, biomarkers can be classified into five different groups: i) diagnostic, ii) prognostic, iii) predictive, iv) stratification and v) risk factor assessment biomarker. Here, “predictive” and “prognostic” biomarkers can be used interchangeably. This is based on the fact that out of a given prognosis, a certain predication can be deviated and *vice versa*.

Diagnostic biomarker

This type of biomarker is used to either rule-in or rule-out a certain disease the marker is associated with, based on its quantitative state. Hence, the detection of this kind of biomarker at a certain abundance level only allows for the diagnosis. Since the biological samples are “screened”, these types of biomarker can also be called “screening biomarker”. Additionally, screening for sets of biomarker for a single disease or biological state guarantees higher precision and substantiates the associated diagnosis. One of the most prominent examples of a diagnostic biomarker is the over-the-counter available pregnancy test. Simply by verifying the existence (at a certain biological availability in the urine) of hCG in a female’s urine, a pregnancy can be indicated.

Prognostic biomarker

Apart from diagnostic biomarkers, which help to determine the answer to a binary question

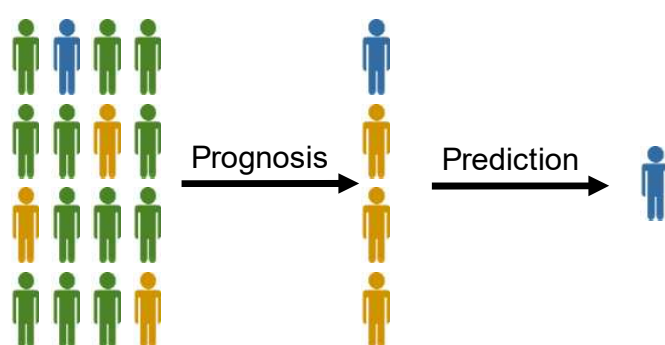


Figure 1.9: Illustration of “prognostic” and “predictive” biomarkers. Figure in style of Brünner (2009).

(healthy or diseased state), prognostic biomarkers provide profound information on the likely course of the present disease in an untreated patient. Prognostic markers can display the risk for metastasis formation and further scattering into adjacent tissue, especially for cancer. Types of breast cancer demonstrating abnormally high levels of *HER2/neu* are

considered highly aggressive and if untreated will easily form metastasis and consequently lead to an early death of the diseased patient.

Predictive biomarker

As already mentioned, the term “prognostic” and “predictive” biomarker can be used interchangeably in certain cases. Nevertheless, “predictive” biomarkers are most frequently used to identify subpopulations of patients for whom a positive response to a given therapy can be predicted. Since these markers are used to predict a successful therapy, these markers are also the first step towards personalized healthcare. A good example for the interchangeability of prognostic and predictive function imposes HER2/*neu* positive breast cancer biopsies. On one hand the HER2/*neu* marker describes the predominant aggressiveness, but on the other hand it predicts a positive patient’s response while applying a “Herceptin” (monoclonal antibody [mAb] against HER2/*neu*) therapy.

Stratification biomarker

The stratification biomarker could be considered a group or set of biomarkers. Its function is defined as classifying or grouping patients based on characteristics of diseases which are in common. Based on the information, the best outcome in terms of prevention and therapy success can be estimated. Hence, stratification biomarkers specifically share informational content with prognostic and predictive biomarkers.

Risk factor assessment

Besides only assessing disease-specific characteristics, biomarkers can also be used to estimate the predisposition of a patient to a particular disease. Additionally, risk factor assessment helps to detect the early onset of diseases and hence can support prevention. However, not only can the disease prevention be supported, but the assessment of symptoms which have already occurred can also be correlated with the early onset of a disease.

1.2.2 Biomarker Discovery

For decades, the discovery of biomarkers have increasingly moved into the scope of biomedical research. Due to technological advancements and high-throughput technologies, comprehensive analysis of genes, transcripts, proteins, metabolites or other disease related molecules is possible nowadays. In this context, MS-based proteomics methods feature a special promise for the discovery of biomarkers, particularly in their ability to identify thousands of proteins and peptides within a biological sample such as tissue, serum/plasma, urine or CSF.

The search for, or the discovery of biomarkers can be generalized into two basic approaches: i) hypothesis-driven or ii) discovery-based. The hypothesis-driven approach uses the existent knowledge of diseases and its underlying biological processes to deviate hypotheses, which have to be proven by the conducted experiments. For example, knowledge about Diabetes mellitus and its related high blood-glucose levels, led to the discovery of glycosylated hemoglobin as biomarker (HbA_{1c}). In contrast, the discovery-based approach focuses on the identification of changes in the abundance of molecules which in turn are closely related to the disease state. In this regard, quantitative shotgun-proteomics imposes a useful and valuable tool as thousands of features can be profiled simultaneously.

Nonetheless, the study design also imposes a crucial step within the whole process of biomarker discovery. In general, biomarkers, or characteristic differences that define the state of a disease can be discovered by comparing a “healthy” state (i.e. control) with a diseased state (i.e. case). However, by simply comparing only a healthy state with a diseased state, a biomarker for only a certain symptom associated with the disease can be found, instead of an actual disease-related biomarker. For example, by only comparing the CSF of patients suffering from Alzheimer’s disease (AD) with the CSF of healthy patients, the biomarker candidate found might only reflect the presence of dementia, which for example, also occurs while suffering from Parkinson’s disease (PD), but not the specific underlying disease. Hence, the selection of appropriate controls for a biomarker discovery study imposes a crucial step in the whole process and has to be clarified upfront. The general workflow can then be divided into four distant sections: i) Discovery, ii) Qualification, iii) Verification and iv) Validation (cf. **Figure 1.10**). Here, during the discovery phase an unbiased approach is applied and hence, thousands of possible candidates are screened. While changing the methodology from an unbiased procedure to a targeted approach, the number of analytes narrows down each further step. Besides narrowing down the number of possible biomarker candidates, the sample size, which in turn represents the future target population, is steadily increased. Finally, only a few candidates will then enter a clinical validation. Only if the biomarker candidate or the panel of biomarkers proof specificity and sensitivity in a representable cohort during the clinical validation, the candidates receive U.S. Food and Drug Administration (FDA) approval and can be used as a commercial tool.

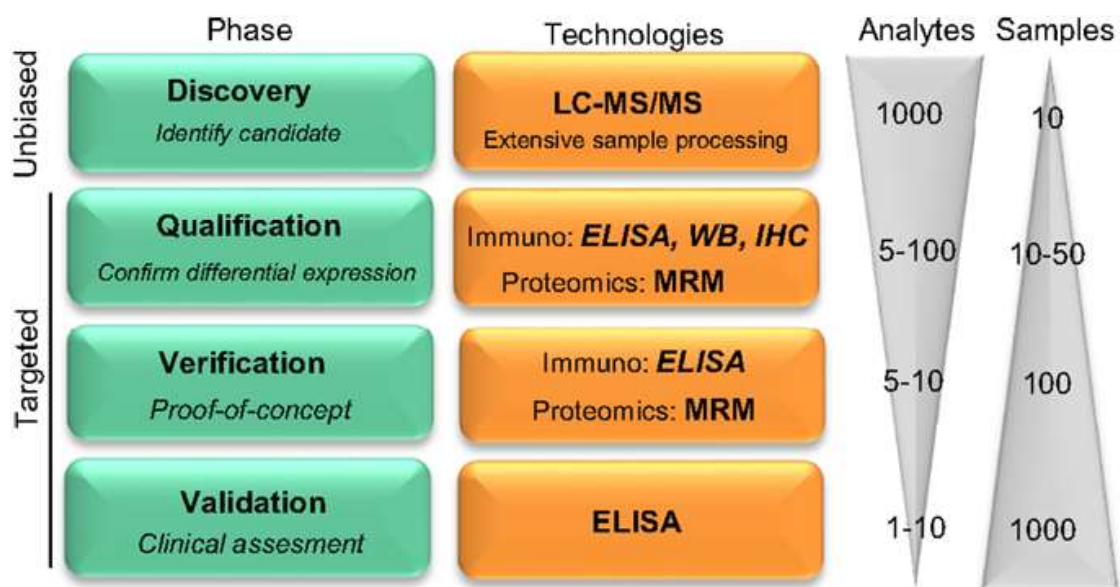


Figure 1.10: Illustration of the proteomic biomarker discovery workflow. The process of biomarker discovery can be divided into four distinct sub-processes: discovery, qualification, verification and validation phase. While the discovery phase has to be performed unbiased, all consecutive steps are performed on a target approach. Besides changing methodology per phase, also, the number of analytes decreases while the number of samples to be analyzed increases. The whole process of biomarker discovery can be summarized as “finding a needle in a haystack”. Figure was adopted from Del Campo *et al.* (2015) ⁶⁴

1.3 Neurological Injuries

The human nervous system can be divided into two functional categories: i) central nervous system (CNS) and ii) peripheral nervous system (PNS). The CNS is represented by the brain and its related cell types as well as the spinal cord, while the PNS describes nerves, which connect parts distant parts of the body with the CNS through their long-reaching axons^{65, 66}. Based on the flow of information, nerves or what is known on a cellular level as neurons, are categorized as either sensory (information from PNS to CNS) or motor neurons (movement stimuli from CNS to PNS). The most recent estimation concluded that the human brain consists of close to 100 billion neuronal cells. This in turn gives rise to the assumption that a typical adult brain features between 100 to 500 trillion synaptic connections. Nevertheless, the function of each present brain cell might be understood, but the way in which millions of cells can cooperate simultaneously still remains an unknown⁶⁷⁻⁶⁹. However, any malicious change or malfunction in this highly complex network will inevitably lead to a functional disruption. As the brain is seen as the control center of the human body, a disruption to its network and intra-connections will consequently lead to a loss of body control functionality. These underlying malfunctions are normally caused by loss of neuronal activity or even neuronal cell loss by induced cell death due to genetic defects, traumatic injury (physical force to the head), toxic influence, infection or simply by aging. Nonetheless, these causes can be generalized as either: i) Acquired Brain Injury (ABI) or ii) Neurodegenerative disorder (ND)⁷⁰⁻⁷².

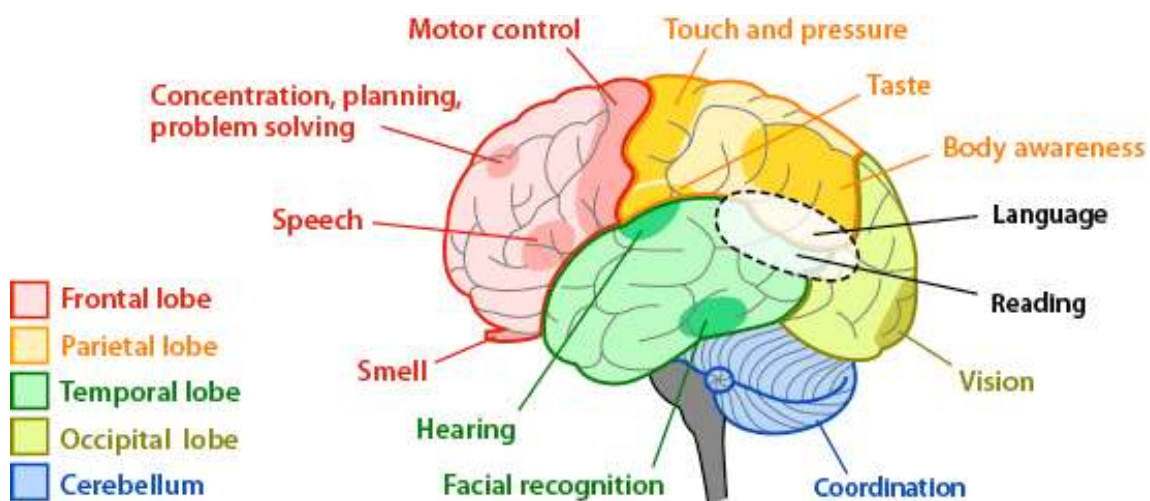


Figure 1.11: The brain and its functional areas. The brain imposes the central control center of the human body. Each bodily function is coordinated in a different section of the brain: frontal lobe (red), parietal lobe (orange), temporal lobe (green), occipital lobe (ocher) and cerebellum (blue). Injuries to the brain will affect the bodily functions which are associated to the region the injury occurs. Figure adopted from <https://askabiologist.asu.edu>.

1.3.1 Acquired Brain Injury

The ABI describes neuronal injuries to the brain by either traumatic events (traumatic brain injury; TBI) or non-traumatic injuries caused by internal or external sources which can be describes as a stroke, tumors, intoxication, substance abuse, hypoxia, encephalopathy or infections. Additionally, the ABI term excludes any congenital or post-natal acquired neurodegenerative disorders (cf. **Figure 1.12**)^{73, 74}.

Since non-traumatic neuronal injuries are mainly related to other clinical pictures and are rather considered as secondary injuries due to other underlying pathological effects or disorders, here, the emphasis lies on TBIs.

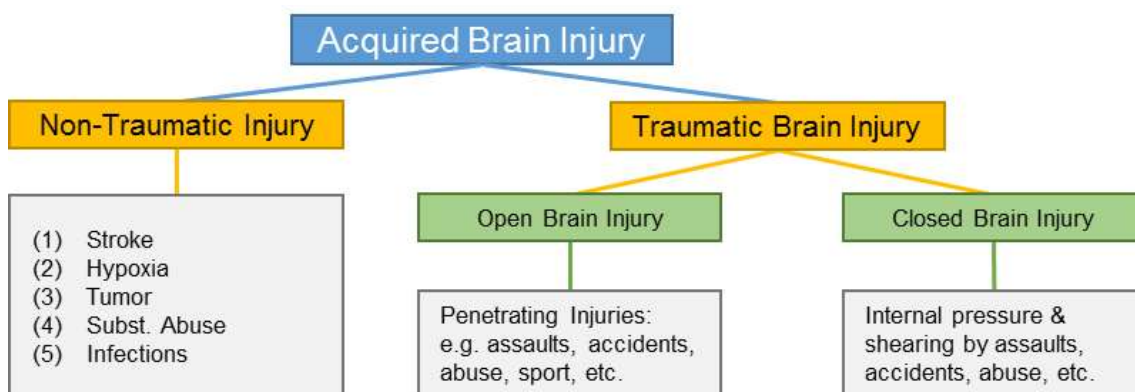


Figure 1.12: Schematic explanation of “Acquired Brain Injury” (ABI). The term “Acquired Brain Injury” describes on the one hand “Non-Traumatic Injuries” which can occur as secondary events due to stroke, hypoxia, brain tumors, substance abuse, infections or encephalopathies. On the other hand, “traumatic brain injuries” which can occur as an “open brain injury” when the skull or brain tissue is actively pierced or as “closed brain injury” as the injury is caused by internal skull pressure or shearing of brain tissue without piercing, are also considered as ABI. Figure after “Brain Injury Association of Massachusetts” (2011; <http://www.biama.org/braininjury>).

In this regard, a closed brain injury is caused by either a sudden and violent blow to the head with an object or by the head hitting an object at a fast pace. Additionally, in certain instances, the object which hits the head might have enough force to pierce the skin and skull and hence primarily damage the brain tissue. In this case, the TBI is called an open brain injury. Nonetheless, TBIs are not only limited to accidental events, but might also occur during sporting activities like soccer, boxing, etc. It is estimated that each year approximately 1.7 million people in the United States of America alone, sustain a TBI⁷⁵. Nonetheless, each TBI is different as symptoms can vary from mild (mTBI), moderate or severe. Each clinical picture or i.e. the occurrence of symptoms consequently depends on the extent of the brain injury. While the actual neuronal injury is mainly caused at the time of impact (e.g. fracture, bleeding, etc.), secondary neuronal injuries also state a significant cause of brain damage (intracranial pressure, brain swelling, etc.). Typically occurring symptoms for TBIs can be found in **Table 1.3**⁷⁶⁻⁷⁹.

Table 1.3: Exemplary list of symptoms for traumatic brain injuries.

Unconsciousness	Confusion	Disorientation	Headache
Nausea	Vomiting	Drowsiness	Constant sleeping
Weakness	Seizures	Pupil dilatation	Depression
Anxiousness	Slurred speech	Coma	Sensitivity to light
Loss of concentration			

Traumatic brain injuries are currently diagnosed by physicians and the combination of patients' reports, general state of health and brain scans (CT or MRI scans). Nonetheless, mTBIs in particular do not show any injury-specific features in the brain scans, as physicians then only rely on the patients' reports and symptoms, which have already occurred. Additionally, physicians have limited options for treating TBIs. Since very little can be done to reverse the initial injury or impact, physicians focus on the stabilization of the patient and the prevention of any further brain injury. Hence, the best treatment for mild traumatic brain injuries lies within the prevention ⁷⁹⁻⁸².

In about half of the moderate or severe cases, patients need operations to repair or remove contusions (bruised brain tissue) and/or the hematomas (ruptured blood vessels) to assure further proper oxygen supply for the remaining and adjacent brain tissue as well as avoiding any further brain damage. Regardless of the treatment success, patients can suffer from disabilities such as cognition impairments (memory, thinking and reasoning), sensory processing disabilities (sight, hear, touch, taste and smell), reduced communication abilities, behavioral or mental health impairments (anxiety, depression, aggression, etc.) according to the severity of the injury. In very serious cases, patients even fall into a persistent vegetative state (PVS) and demonstrate no neurological function or activity ⁸³⁻⁸⁶.

1.3.2 Neurodegenerative Disorders

The term "neurodegenerative disorders" is an umbrella term for a range of conditions and processes which primarily affect neuronal cells in the human brain. In contrast to ABIs, NDs can be of a congenital nature, be caused by aging (mitochondrial DNA mutation or oxidative stress) or even occur as a consequential event of ABIs. Additionally, NDs are considered incurable which results in an inexorably progressive neuronal degeneration or cell death of neurons. It appears that many disorders can be related to each other on a sub-cellular level ⁸⁷⁻⁸⁹. These similarities give hope to be able to cure or treat different diseases simultaneously with the same therapy. Neurodegeneration is generally found in many different levels of the neuronal interconnection which ranges from molecular to systemic. In the following, the two most common neurodegenerative diseases are briefly described.

Table 1.4: Chart of in ND involved mechanisms.

Mechanism	Target molecule	Disease
Genetic	Polyglutamine (PolyQ)	Huntington's disease (HD) ⁹⁰
Protein misfolding	α -Synuclein	Parkinson's disease (PD) ^{91, 92}
	Tau	Alzheimer's disease (AD) ⁹³
	β -Amyloid	AD ⁹⁴
	Prion	Transmissible spongiforme Enzephalopathie (TSE) ⁹⁵
Protein degradation	β -Amyloid	AD, PD, HD
Membrane damage	α -Synuclein	PD
Mitochond. dysfunction	Oxidative species	(multiple)
Pathogenic cell death	Programmed cell death	(multiple)

Alzheimer's disease

Morbus Alzheimer or Alzheimer's disease (AD) describes the most common neurodegenerative disease with a prevalence of 60 to 70% of dementia cases. Early symptoms include short-term memory loss and dementia like symptoms and with the progression of the disease, the following symptoms will also occur: problems with language, disorientation, mood swings, loss of motivation, no self-care and behavioral impairments ⁹⁶. The progression of the disease depends on the patient's conditions but the loss of bodily functions will consequently lead to death within three to nine years after diagnosis, yet the actual cause of the disease remains unknown ⁹⁷. Nonetheless, on the molecular level, AD has been classified as a protein misfolding disease (proteopathy). Here, misfolded β -Amyloid (A_{β}) and/or tau protein accumulate in the brain and cause neuron degeneration. The A_{β} accumulation is caused by the aggregation of A_{β} peptides with a length of 39 to 43 amino acids. Here, the A_{β} peptides originate from the "amyloid precursor protein" (APP) which in turn is a transmembrane protein of neurons and triggers neuronal growth, survival and post-injury repairs, under its regular biological condition. In AD the two proteases γ -secretase and β -secretase are jointly digesting the APP protein while one of the proteolysis products, A_{β} , in particular, forms the characteristic senile plaques ⁹⁸⁻¹⁰¹.

Nonetheless, AD can also be considered as taupathy. Here, the microtubule-associated protein "tau" is hyper-phosphorylated which leads to pairing with other threads and the formation of neurofibrillary tangles which then disintegrates the neuron's transport system. In a healthy state neuron, tau is singly phosphorylated and stabilizes the formation of microtubules, which act like transport paths for nutrition and molecules from the neuron cell body to the end of the axon ¹⁰².

Table 1.5: Exemplary list of neurodegenerative disease classifications. List adopted from Mackenzie *et al.* (2010)⁷²

Classification	Associated Disease
Tauopathy	Alzheimer's disease
	Progressive supranuclear palsy (PSP)
	Corticobasal degeneration (CBD)
Synucleinopathy	Parkinson's disease
	Dementia with Lewy bodies (LBD)
Trinucleotide repeat disorder	Chorea Huntington (HD)
	Friedreich's ataxia
Motor neuron disorders	Amyotrophic lateral sclerosis (ALS)
	Spinal muscular atrophy (SMA)
Transmissible spongiform encephalopathy	Creutzfeldt–Jakob disease (CJD)
	Gerstmann–Sträussler–Scheinker syndrome

Parkinson's disease

Morbus Parkinson or Parkinson's Disease (PD) represents the second most common neurodegenerative disease with a prevalence that ranges between 300,000 and 400,000 people in Germany alone¹⁰³. In the early stages, symptoms are commonly describes as shaking, rigidity, slowness of movement and difficulty with walking. Once the disease progresses, dementia can also occur as an additional symptom. Additionally, the aforementioned symptoms can be accompanied by sensory, sleep and emotional problems^{104, 105}.

On a molecular level, PD is described as proteopathy and is hence comparable to some extend to the molecular processes involved in AD: one hypothesis indicates the abnormal accumulation of α -Synuclein which in turn is bound to ubiquitin. The insoluble protein aggregates in the cytosol of neurons and forms inclusion bodies which are called "Lewy Bodies". These Lewy bodies are consequently unable to be directed correctly to the proteasome. However, the formation of Lewy bodies might not be necessarily causing cell death but can provide protective function to the affected neurons. Recent research revealed that the death of dopaminergic neurons by α -Synuclein is caused by a defect in the transport machinery of proteins between the endoplasmic reticulum (ER) and the Golgi apparatus. It is further hypothesized that proteosomal and lysosomal system dysfunctions as well as reduced mitochondrial activity can cause cell death in PD. Nonetheless, the mechanism is not yet fully understood¹⁰⁶⁻¹⁰⁸.

2. MStern Blotting

2.1 Introduction

Mass spectrometry (MS)-based proteomics is moving increasingly into the translational and clinical research arena, where robust and efficient sample processing is of particular importance. The conventional sample processing methods in proteomics, namely SDS-PAGE- or in-solution-based sample processing, are slow and laborious, and thus do not easily provide the reproducibility and throughput to meet current demands. A paradigm shift was the introduction of filter-aided sample processing method (FASP), which were initially described by Manza et al. (2005)¹⁰⁹ and then fully realized in practice by Wisniewski et al. (2009)¹¹⁰. These filter-aided methods make use of ultrafiltration membranes with molecular weight cut offs (MWCO) in the 10 to 30 kDa range to efficiently remove small molecules and salts, and to capture denatured proteins on a cellulose filter even if the molecular weight of the protein is much smaller than the nominal MWCO of the ultrafiltration membrane. Thus, the denaturation step is crucial to ensure that proteins much smaller than the nominal MWCO are efficiently retained by e.g. a 10 kDa MWCO filter.

In translational and clinical proteomics, which normally include large cohorts, the multi titer-well plate is the preferred format for sample processing and storage. Although the application of FASP in the 96-wellplate format has been described^{111, 112}, the major limitation of FASP in the 96-well plate is the much slower speed at which the 96-well plates have to be centrifuged, while a single ultrafiltration unit withstands up to 14,000×g, the 96-well plate format can only be centrifuged at g-forces of up to 2,200×g. This significantly lower g-force for 96-well plates results in a slow liquid transfer, which in turn considerably prolongs the required centrifugation times to hours instead of tens of minutes for the three to four necessary centrifugation steps i) for the initial loading, reduction and alkylation, ii) for the different washing steps, and iii) for the elution¹¹¹.

Independent of the format FASP is performed in, the conventional FASP also requires relatively large volumes of high salt concentration for efficient elution of the tryptic peptides. Hence, reversed phase-based desalting of the samples is a prerequisite for subsequent LC/MS experiments. Apart from prolonging the entire FASP procedure, the numerous additional handling steps are also potentially associated with peptide losses¹¹³.

In this study, a description is given of novel sample processing workflow for MS-based proteomics that utilizes the strengths of filter-aided sample processing methods and at the same time overcomes their major limitations, without compromising the results, i.e. significantly reducing the number of identified peptides and/or proteins. The result is a significantly improved throughput as 96 samples (or multiples there-of) can be completely processed within a single workday.

2.2 Experimental Procedures

2.2.1 Cell Culture

Human cervical cancer cells (HeLa) were propagated in Dulbecco's modified Eagle's medium (DMEM; 11965; Life technologies). Upon achieving 85-90% confluence, the growth media was aspirated and the cells were washed three times with 5ml ice-cold PBS. One ml of modified RIPA buffer (150 mM NaCl, 50 mM Tris/HCl pH 7.4, 1% NP-40, 0.1% sodium deoxycholate, 1 mM EDTA) supplemented with 1x Roche Complete protease inhibitors, was added to each plate of cells and incubated for 30 min on ice. Cells were scraped with a cell scraper, collected in Eppendorf tubes and vortexed for 1 min. Cellular debris and other particulate matter was pelleted by centrifugation at 20,000×g at 4°C; the supernatant was recovered for further use.

2.2.2 Urine Collection for Ovarian Cyst Biomarker Study

Urine samples were collected from consenting patients visiting the Emergency Department at Boston Children's Hospital. The study was reviewed and approved by Boston Children's Hospital's Internal Review Board (Protocol Number X06-10-0493).

2.2.3 Protein Concentration Determination

Protein concentration was determined by using the Bradford Assay ¹¹⁴ (Bio-Rad DC™ Protein Assay) following the manufacturer's protocol. The standard curve was established using a stock solution of 20 mg/ml bovine serum albumin (BSA) and final concentrations of 0.25 mg/ml, 0.5 mg/ml, 1 mg/ml, 1.5 mg/ml and 2.0 mg/ml. After incubation at room temperature (RT) the final measurement was performed in a micro plate spectrophotometer (Bio-Rad Model 680) at a wavelength of 595 nm.

2.2.4 MStern Blot

Undiluted neat urine (150 µl, i.e. ~15 µg of protein) was added to a mixture of 150 µg urea and 30 µl dithiothreitol (DTT) (100 mM in 1 M Tris/HCl pH 8.5). Diluted HeLa cell lysates (10 µg in 100 µl 50 mM ammonium bicarbonate (ABC)) or neat CSF (10 µl, i.e. ~10 µg of protein) was added to 100 µg urea and 20 µl DTT. The resulting solution was incubated for 20 min at 27°C and 1100 rpm in a thermo mixer. Reduced cysteine side chains were alkylated with 50 mM iodoacetamide (IAA; final concentration) and incubation for 20 min in the dark at 27°C and 750 rpm.

The hydrophobic PVDF membrane in a 96-well plate format (MSIPS4510, Millipore) was pre-wetted with 150 µl of 70% ethanol and equilibrated with 300 µl urea supernatant (~8.3 M urea). These and all subsequent liquid transfers were carried out using a fitted 96-well microplate vacuum manifold (MAVM0960R, Millipore).

Each sample was transferred three times through the PVDF membrane by applying a vacuum. The addition of Ca²⁺ was also tested, which had been described as beneficial for the protein binding onto PVDF membranes ¹¹⁵, but were not able to notice a benefit. After protein adsorption of the proteins onto the membrane, it was washed twice with 50 mM ABC.

Protein digestion was performed with sequencing grade trypsin (V5111, Promega) at a nominal enzyme to substrate ratio of 1:15. To this end, 100 μ l digestion buffer (5% acetonitrile (ACN; v/v), 50 mM ABC and trypsin) were added to each well.

After incubation for 2 hours at 37°C in a humidified incubator, the remaining digestion buffer was evacuated. Resulting peptides were eluted twice with 150 μ l of 40% ACN (v/v)/0.1% (v/v) formic acid (FA) each. Upon pooling, the peptide solutions were dried in a vacuum concentrator. Lyophilized samples were stored at -20°C until further analysis.

Note: different batches of filter-plates were tested and no batch dependent performance differences were observed.

2.2.5 Filter assisted sample preparation

The filter assisted sample preparation method was carried out as previously described¹¹⁰. In short, proteins were first denatured and reduced by adding 100 μ l sample to 100 μ g urea supplemented with 20 μ l DTT. For the different sample types, namely urine, CSF and HeLa lysate a nominal protein content of 15 μ g, 10 μ g and 10 μ g, respectively were used for analysis. After alkylation of reduced cysteine side chains with 50 mM IAA (final concentration), denatured proteins were captured on a 10 kDa MWCO spin filter (MRCPRT010, Millipore) and washed twice with 50 mM ABC. Protein digestion was performed with sequencing grade trypsin (V5111, Promega) at a nominal enzyme to substrate ratio of 1:50. After incubation over night with 100 μ l digestion buffer (trypsin in 50 mM ABC), resulting peptides were eluted with 300 μ l 0.5 M NaCl. Peptide elutes were desalted with reversed phase-based TARGA C-18 spin tips (SEMSS18R, Nest Group) prior to LC-MS/MS analysis. Lyophilized samples were stored at -20°C for further analysis.

2.2.6 LC-MS/MS Analysis

Peptides were reconstituted in loading buffer (5% ACN (v/v), 5% FA (v/v)). LC-MS/MS analysis was performed on a microfluidic chip system (EK425) coupled to a TripleToF 5600+ mass spectrometer (both Sciex). Tryptic digests (~1 μ g) were loaded onto a trap column (ReproSil-Pur C₁₈-AQ, 200 μ m x 0.5 mm, 3 μ m) and subsequently separated on a ReproSil-Pur C₁₈-AQ analytical column chip (75 μ m x 15 cm, 3 μ m) at a flow rate of 300 nl/min. A linear gradient from 95% to 65% buffer A (0.2% FA in HPLC water; buffer B: 0.2% FA in ACN) within 60 min was applied. Samples were ionized applying 2.3 kV to the spray emitter. Analysis was carried out in a data-dependent mode. Survey MS₁ scans were acquired for 200 ms. The quadrupole resolution was set to 'UNIT' for MS₂ experiments, which were acquired for 50 ms in 'high sensitivity' mode. Following switch criteria were used: charge: 2+ to 4+; minimum intensity: 100 counts per second (cps). Up to 35 ions were selected for fragmentation after each survey scan. Dynamic exclusion was set to 17 s.

Samples for the ovarian cyst biomarker study were analyzed using the same LC-chip system (Sciex, trapping column: 200 μ m x 0.5 mm ReproSil-Pur C₁₈-AQ 3 μ m; analytical column: 75 μ m x 15 cm ReproSil-Pur C₁₈-AQ 3 μ m) coupled online to a Q Exactive mass spectrometer (Thermo

Fisher Scientific). Peptides (4 µl of digest) were separated by a linear gradient from 93% buffer A (0.2% FA in HPLC water) / 7% buffer B (0.2% FA in ACN) to 75% buffer A / 25% buffer B within 75 min. The MS was operated in data-dependent TOP10 mode with the following settings: resolution for MS1 scan 70,000 @ 200 Th; lock mass: 445.120025 Th; resolution for MS2 scan 17,500 @ 200 Th; isolation width 1.6 m/z; Normalized Collision Energy (NCE) 27; underfill ratio 1%; charge state exclusion: unassigned, 1, >6; dynamic exclusion 30 s.

2.2.7 Data Analysis

Acquired MS raw files (WIFF) were analyzed using ProteinPilot (version 4.5.1, revision 2768; Paragon™ Algorithm 4.5.1.0, 2765; Sciex) using the human UniProtKB database (Homo sapiens, ~20,300 sequences, version 06-2014). The 'thorough' search mode was used. The digestive enzyme was set to trypsin. It is noteworthy that ProteinPilot does not require the definition of an allowable number of missed cleavages, modifications or mass tolerances. Commonly occurring laboratory contamination protein sequences (cRAP, version 2012.01.01) were added to the database.

For the label free quantification, the precursor intensities was extracted from the *Spectrum Summaries* generated by ProteinPilot. Intensity-based absolute protein quantitation (iBAQ) ¹¹⁶ for dynamic range analysis, MaxQuant ¹¹⁷ (version 1.5.1) was used. In short, the acquired WIFF files were loaded into MaxQuant and searched against the human UniProtKB database (Homo sapiens, ~20,300 sequences, version 06-2014). For quantification, the 'iBAQ' and 'label-free quantification' (LFQ) were selected. Variable Modifications were set to: oxidation(M), deamidation(NQ) and phospho(STY). For fixed modifications only carbamidomethyl(C) was used. Peptides that showed the aforementioned modifications were allowed for quantification, however counter peptides were not rejected. As digesting enzyme 'trypsin' was specified. Otherwise default settings (by MaxQuant) were used for the analysis: trypsin with up to two missed cleavages; mass tolerance for the first search: 0.07 Da; main search: 0.006 Da

For Gene Ontology (GO) annotations-specific isoforms were merged to their parent protein. Samples related to the ovarian cyst biomarker study have been analyzed within MaxQuant 1.5.2.8 and searched against the human UniProtKB database (Homo sapiens, ~20,300 sequences, version 10-2014). The following settings were applied: trypsin with up to two missed cleavages; mass tolerance for the first search: 20 ppm; main search: 4.5 ppm; fixed modification: carbamidomethyl (C); variable modification: oxidation (M), otherwise default settings were used. The iBAQ approach ¹¹⁶ was employed for quantification. Prior to the statistical analysis, samples with less than 200 identified proteins were removed from the dataset (5/89 samples, i.e. 5.6% of the samples). Subsequently, a two-tailed student's t-test was applied to identify the significantly changed proteins. Bonferroni correction was applied to account for multiple testing.

GO annotations were established by FunRich (<http://www.funrich.org>). Venn diagrams were generated using the available online tool, Venny (<http://bioinfogp.cnb.csic.es/tools/venny/>). For the calculation of chemical and physical properties an in-house R-script (v3.15) as well as the online tool "GRAVY Calculator" (<http://www.gravy-calculator.de>) were used. Boxplots and Receiver-Operating Characteristics (ROC) were generated in SPSS v21 (IBM).

All mass spectrometry proteomics data were deposited to the ProteomeXchange Consortium ¹¹⁸ via the PRIDE partner repository: Urinary biomarker study (Ident: PXD002549) and MStern Blot vs. FASP comparison (Ident: PXD002572). Annotated spectra for the biomarker study are also accessible through MS Viewer ¹¹⁹ (SearchKey: "sd9zoqvsoo").

2.3 Results and Discussion

2.3.1 FASP vs. MStern Blot

Filter-based sample processing, in particular, FASP, have established themselves as a widely used proteomics sample processing/digestion method due to their improved sensitivity and throughput when compared to gel-based digestion methods, and due to their improved compatibility with salts, detergents and other denaturants in comparison to conventional solution-based processing methods. Despite a multitude of advantages, FASP or FASP-like methods have the drawback of not being readily compatible with 96well plate formats because of the small pore size of the cellulose-based ultrafiltration membranes. This gives rise to very long centrifugation times when used in the 96-well plate format due to the limited g-force 96-well plates can withstand: $2,200 \times g$ for 96-well plates vs. $14,000 \times g$ for individual ultrafiltration units. Cellulose ultrafiltration

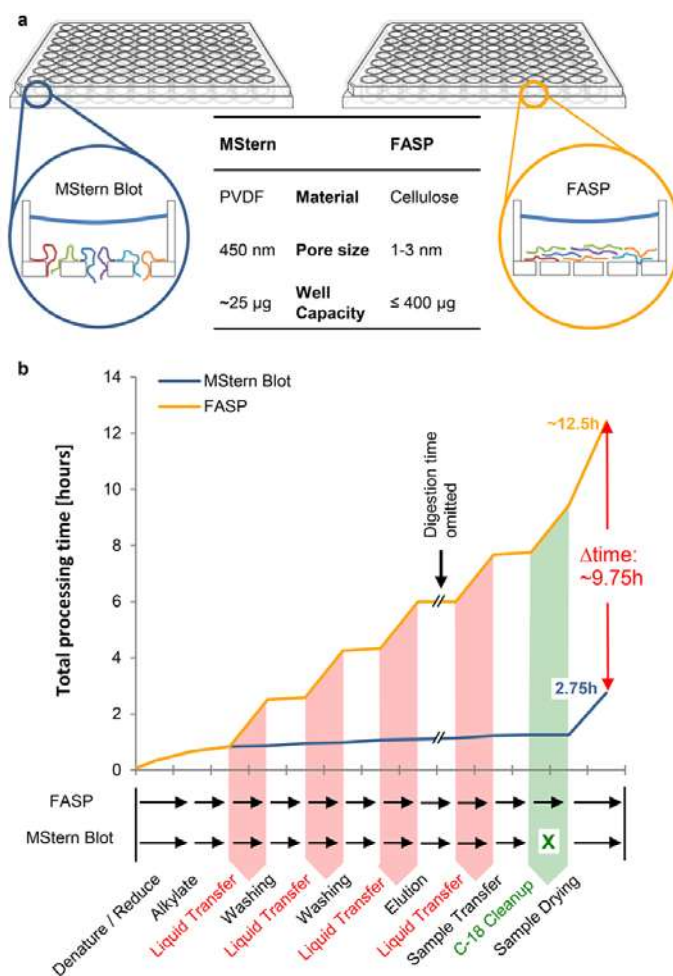


Figure 2.1: FASP vs. MStern Blot. **a)** Comparison of the physical properties of the ultrafiltration membrane used for FASP and the membrane used for MStern Blot. FASP uses physical retention whilst in MStern blotting proteins are adsorbed onto the hydrophobic membrane surface. **b)** Time advantage of MStern blotting (blue curve) vs. FASP (yellow curve) without considering potentially different digestion times. Major timesavers are the fast liquid transfer steps (1 min vs. 100 min; red) and the omission of any desalting (green).

membranes that are used for the FASP approach feature a pore size of 1-3 nm (10 to 30 kDa MWCO) giving rise to this significant flow restriction.

It was reasoned that switching to large pore hydrophobic PVDF membranes, which have a proven high affinity for proteins that is utilized in Western blots on a daily basis, would overcome the problem of slow liquid transfer through the membrane. These hydrophobic PVDF membranes are used for sterilization and/or filtration feature pores in the size range of 220 to 450 nm (*cf.* **Figure 2.1a**), which is 100 times larger than the pores found in the 10 or 30 kDa MWCO ultrafiltration cellulose membranes used for e.g. FASP. The capacity of PVDF-based protein processing is theoretically in the $25 \mu\text{g}/\text{well}$ range ($100 \mu\text{g}/\text{cm}^2$ ¹²⁰), which is lower than in cellulose membrane-based filtration units. Given the sensitivities of current LC/MS systems, this amount is still plentiful

where rarely more than 1 µg is injected per analysis run. The main difference between cellulose membrane and PVDF for membrane-based proteomics sample processing, is the mode of protein retention: while ultrafiltration cellulose membranes retain the proteins by size-based filtration, PVDF retains the proteins by efficiently adsorbing them onto its large hydrophobic surface, which is widely exploited in e.g. Western blotting, giving rise to the name suggested in this study, 'MStern blotting'.

After establishing an initial protocol for MStern blotting, it was evaluated whether the theoretical processing speed advantage can be realized in practice. Indeed, the use of low-grade (e.g. house) vacuum was sufficient for the liquid transfer through the PVDF membrane in 10 seconds to 2 minutes (depending on the numbers of wells used). In contrast, the centrifugation-based liquid transfer through ultrafiltration cellulose membranes in the 96-well plate format, which requires swinging-bucket rotors, requires 1 to 2 hours^{111, 112} for each liquid transfer step. In summary, the fast liquid transfers through the PVDF membrane can be realized in practice and is the major source of time-savings for the MStern blotting sample processing in comparison to FASP (*cf.* **Figure 2.1b**).

Besides a faster liquid transfer through the membrane, further time-savings are realized by the post digestion peptide elution, which uses a simple mixture of acetonitrile and formic acid instead of concentrated salt solutions as in the case of FASP. While the residual amount of ammonium bicarbonate salts are further reduced by the subsequent quick vacuum centrifugation, such that the samples are ready for LC/MS analysis once they have been evaporated to dryness. In contrast, FASP requires a prolonged vacuum centrifugation and subsequent lengthy and expensive reversed phase-based desalting of the digests. Together with the faster liquid transfers, all time savings add up to more than 9.5 hours when processing samples with MStern blotting instead of FASP. In addition, the use of vacuum manifolds also allows for easier automation when compared to FASP, which requires centrifugation.

2.3.2 Performance of MStern Blot

After establishing that using hydrophobic PVDF for membrane-based proteomic sample processing instead of hydrophilic regenerated cellulose as in the case of FASP results in significant time-savings, the compatibility of adsorption of complex protein mixtures with tryptic digestion was investigated. The digestion of individual proteins adsorbed onto PVDF membranes has been described before¹²¹⁻¹²⁴. However, these previous publications used electroblotting to transfer proteins out of an SDS-PAGE onto the PVDF membrane prior to proteolysis and mass spectrometric analysis. Thus, it was not evident that proteins from complex protein mixtures could be quickly adsorbed onto PVDF membrane by forcing dilute protein solutions through the membrane within seconds or minutes using vacuum instead of electroblotting which operates in the hour timescale¹²⁴. To test whether fast adsorption of proteins to hydrophobic PVDF is a compatible with proteomic studies on complex protein mixtures, three different types of samples were used that are either considered a good source for biomarker (neat urine and neat cerebrospinal fluid (CSF)) as well as a highly complex unfractionated whole (HeLa) cell lysate.

The initial digestion optimization resulted in conditions which match or exceed the performance of FASP; thus, it was expected that a more thorough optimization would provide even better results (for details, see **2.2 Experimental Procedures**). Four aliquots were processed for each sample type and analyzed tryptic digests of the different samples types by LC-MS/MS using a 1-hour gradient. These analyses identify 497 ± 58 , $2,733 \pm 160$, and 676 ± 143 proteins from neat CSF, HeLa cell lysate and neat urine, respectively (cf. **Figure 2.2a**). The FASP-based processing of 4 aliquots of the same samples resulted in 561 ± 40 , $2,473 \pm 89$, and 622 ± 133 proteins for neat CSF, HeLa cell lysate and neat urine, respectively. Also the dynamic ranges of the identified proteins as determined using the iBAQ method¹¹⁶ was similar for both sample processing methods: ~ 5 orders of magnitude of the two neat body fluids and ~ 6 orders of magnitude for the HeLa cell lysate (cf. **Figure 2.2b**). These numbers clearly showed that the MStern blotting approach gives protein identification rates at least as good as FASP, irrespective of the nature and complexity of the sample.

Next, the loading capacity of the PVDF membrane was tested. To this end, 5, 10, 15 and 30 μg of HeLa cell lysate was loaded into individual wells of the 96-well plate with the hydrophobic PVDF

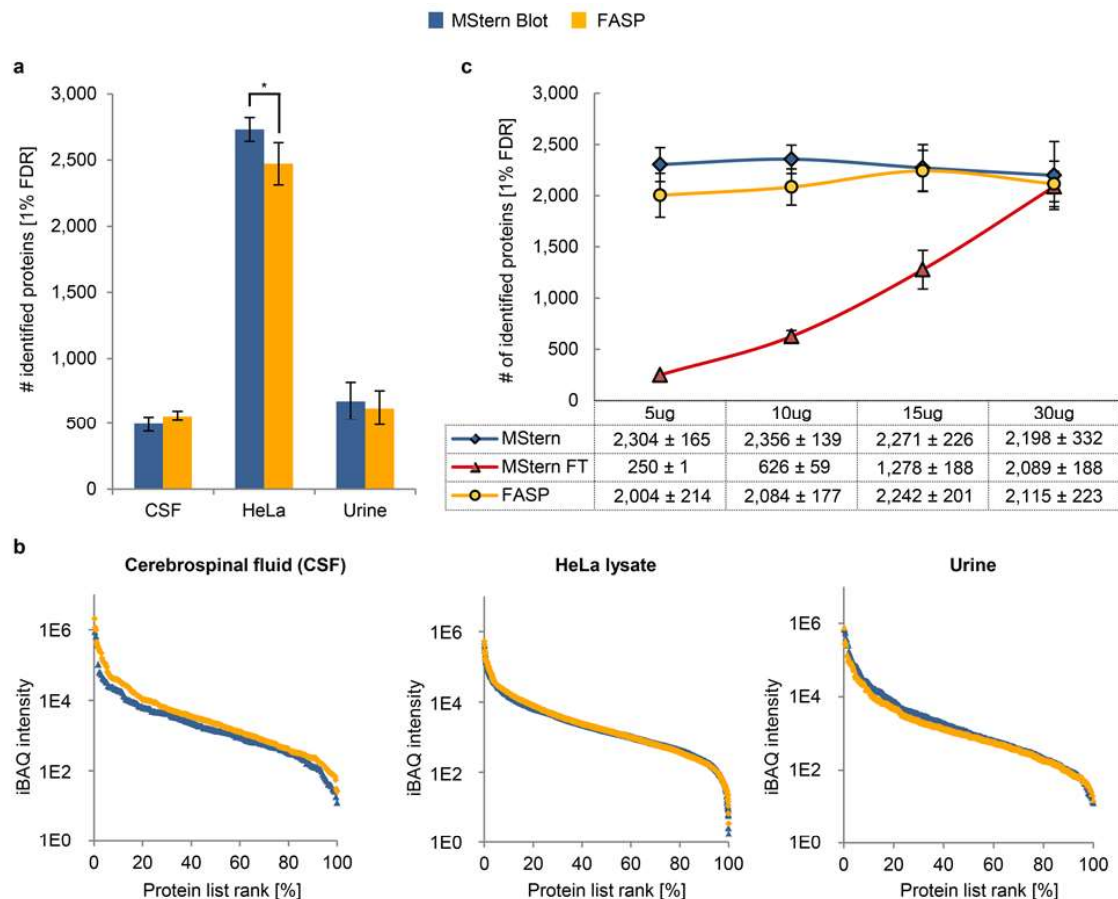


Figure 2.2: Performance Comparison MStern Blot vs. FASP. **a)** Comparison of proteins identified from CSF, HeLa lysate and urine after loading approx. 10 μg , 10 μg , and 15 μg , respectively. Each sample type was processed in quadruplicate. **b)** Testing the loading capacity of the PVDF membrane used for MStern blotting based on proteins identified adsorbed by the PVDF membrane (i.e. MStern blotting, blue curve) and the respective flow through processed by FASP (red curve), in comparison to standard FASP of the same sample (yellow curve). A HeLa lysate was used. Values shown demonstrate average protein identifications. **c)** Comparison of the dynamic range in three different biological samples (CSF, HeLa lysate and urine); MaxQuant-based iBAQ intensities are marked blue (MStern blotting) and yellow (FASP).

membrane at the bottom of each well. The flow-throughs of the loading and washing solutions were collected and subsequently processed using FASP. In parallel, identical amounts of protein were directly processed with FASP (*cf.* **Figure 2.2c**).

In summary, FASP and MStern blotting resulted in a similar number of identified proteins (while MStern consistently identified more than FASP as already shown in **Figure 2.2a**), irrespective of the amount of protein processed. In contrast, the number of proteins identified in the flow through of the MStern blot-based processing steadily increases such that at a nominal loading of 30 μ g, as many proteins are identified in the flow-through as in adsorbed fraction. Based on these numbers, it was concluded that the optimal loading capacity is in the 10 to 15 μ g per well, plenty of material for today's LC/MS instrumentation. These amounts of proteins correspond to e.g. 150 μ l of urine or 15 μ l of CSF, which are easily available from primary body fluids. Thus, additional experiments to test the lower loading limits of MStern blotting were not performed.

2.3.3 Detecting Biases in Proteins Identified in Method-Specific Samples

Since MStern blotting and FASP have very different modes of retention, it was reasoned that both methods would exhibit different preferences for protein identification. To look into this issue, the identification overlap was compared from the combined search results of the four MStern blot and four FASP preparations of neat urine, HeLa cell lysate and neat CSF, which were used to generate **Figure 2.2a**. The Venn diagrams clearly show that 2/3 to 3/4 of the identified proteins are shared between the MStern blot and the FASP method, while 1/4 to 1/3 of the proteins are unique to either MStern blotting or FASP (*cf.* **Figure 2.3a-c**). The commonalities and differences at the peptide level were also compared. Here, specific peptides were in the 50 to 60% range such that only as little as 40% of the observed peptides were in common.

For the subsequent GO annotation of the method-specific proteins the funrich.org tool was used, which uses more broadly defined ontologies to make comparisons more generalizable. **Figure 2.3** shows the results of these comparative protein localizations, where only the 12 most populated GO terms are listed. For neat urine and HeLa cell lysates, only minor differences are observable for the major GO terms. Slightly bigger differences are observable for the neat CSF, such as MStern blotting's biases against plasma membrane and extracellular proteins, and preferences nucleolar, mitochondrial and/or cytosolic proteins. However, a clear trend is not obvious.

2.3.4 Physico-Chemical Properties

To better understand the process-specific differences in the identified proteins and peptides, a closer look at the physicochemical properties of the unique and shared proteins and peptides was taken (*cf.* **Figure 2.4** – the graphs for the HeLa cell lysates are shown; the graphs for CSF and neat urine can be found in the Appendix **8.1 MStern Blotting**). In particular, the molecular weight, the pI and the hydrophobicity/GRAVY score was compared. For reference purposes, the theoretical distribution of all human proteins found in SwissProt was also shown. Comparing the plots for the proteins (*cf.* **Figure 2.4a**), it is apparent that FASP is biasing in favor of small (low molecular weight), charged (higher and lower pI) and more hydrophilic (lower GRAVY score) proteins. In contrast, MStern blot has a slight preference for larger and less charged proteins. These observed

dissimilarities match the differences in the binding modes used for the two sample processing strategies.

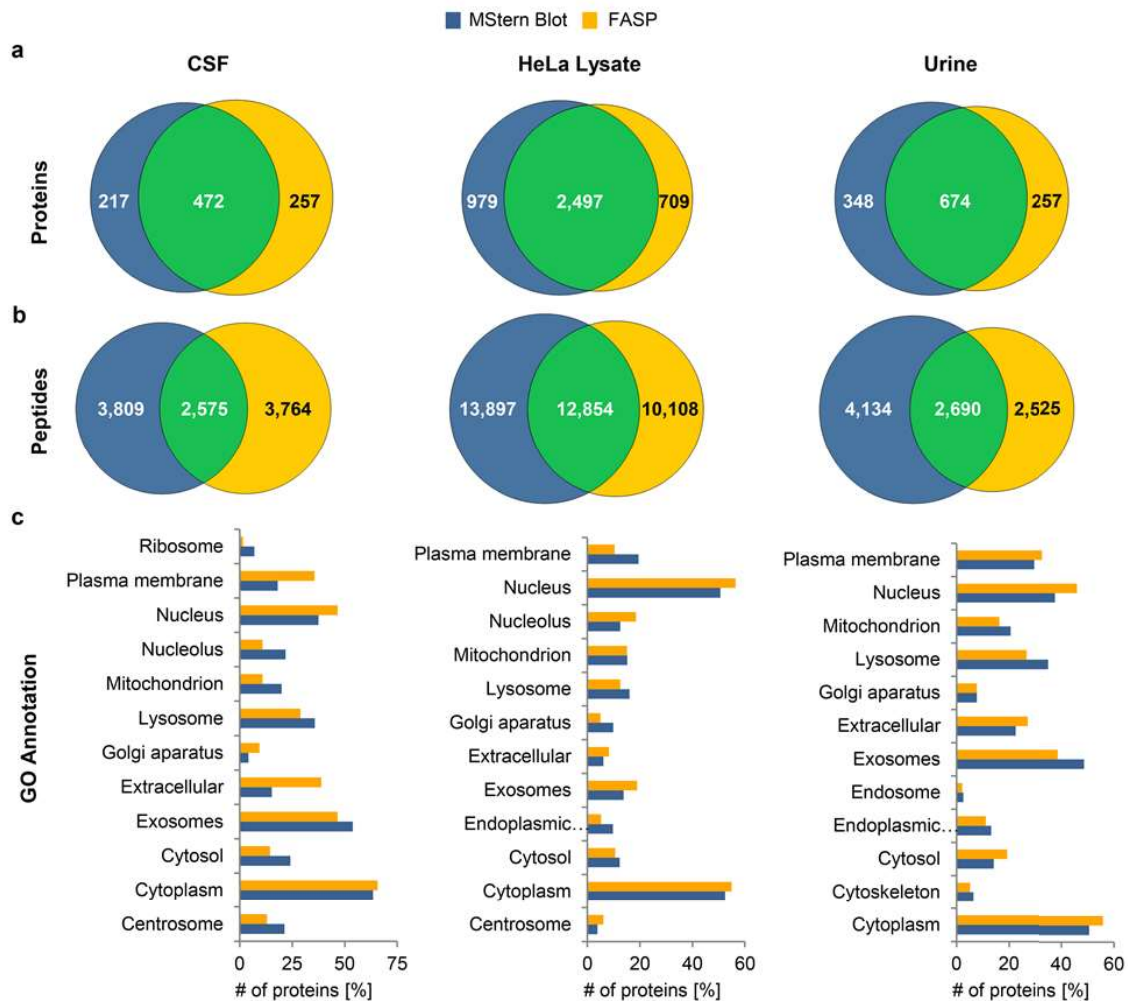


Figure 2.3: Comparison of the Properties of Identified Proteins. Venn diagram of the proteins and peptides identified from CSF (a), HeLa lysate (b) and urine (c). On the bottom, GO annotations (cellular compartment) of the method specific proteins, namely MStern blotting (blue) or FASP (yellow).

Comparing the physicochemical properties of the peptides (cf. **Figure 2.4b**) identified a major shift of the molecular weight of the MStern blot specific peptides. The MStern blot-specific peptides also showed a shift away from lower pI-values in favor of higher pI values above a pI of 6.8, and a minor shift towards less hydrophilic peptides. The latter was unexpected, as larger peptides are generally assumed to more hydrophobic. For reference purposes, the distributions for the theoretical tryptic peptides from all human proteins found in SwissProt are shown, assuming no missed cleavage or two missed cleavages.

Investigating the major shift in the molecular weight distributions of the observed process-specific peptides revealed an increase in peptides with missed cleavages from 12.5% to 37.4% for the MStern blot vs. FASP. Attempts to modulate the degree of missed cleavages by varying the content of organic solvent ¹²⁵ and/or the digestion time had only minor effects, which might indicate that the adsorption of the proteins seem to interfere with the trypsinization.

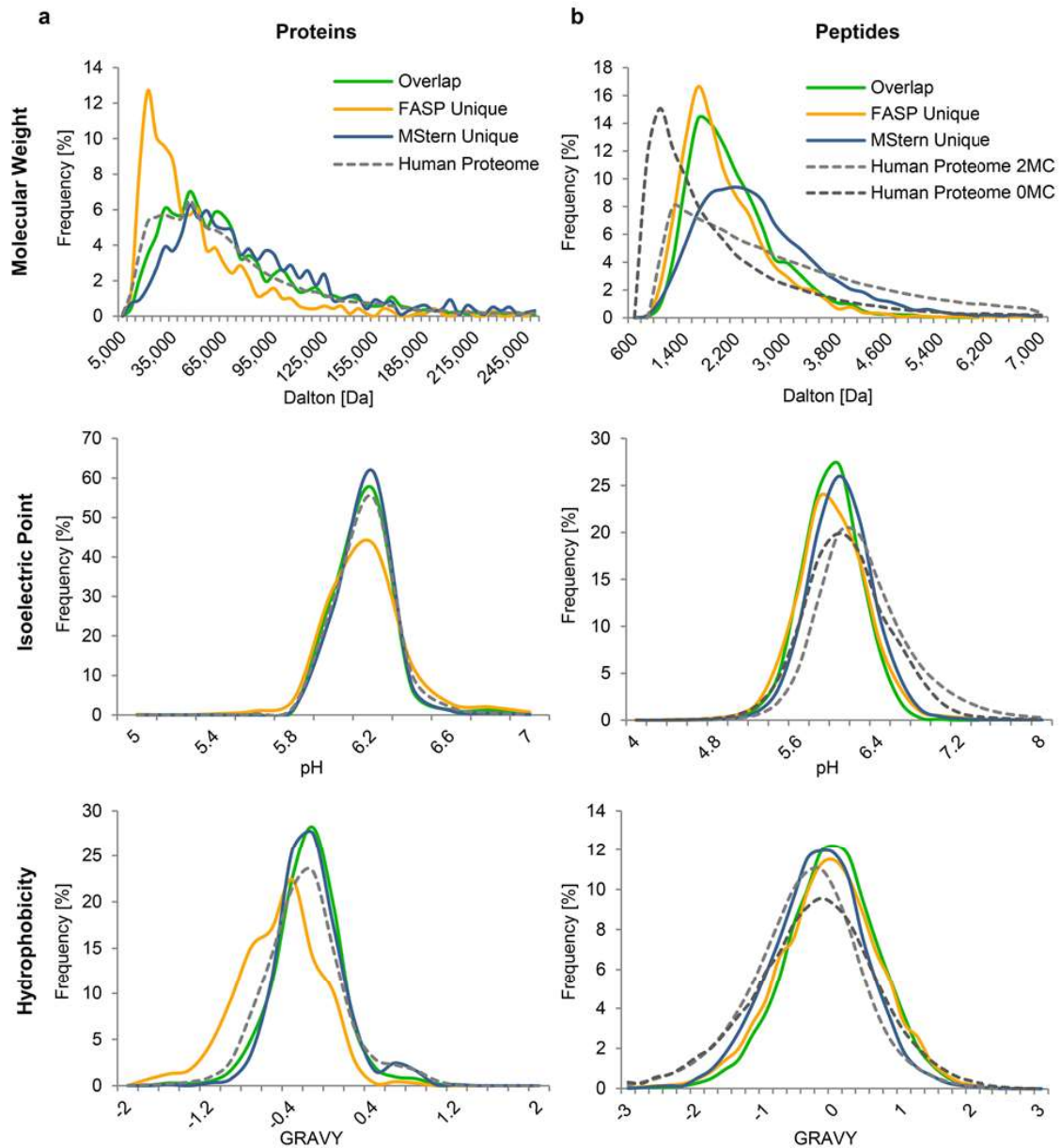


Figure 2.4: Physical-chemical Properties. **a)** Comparison of three different properties: Molecular Weight (top), isoelectric pH (middle) and GRAVY score (bottom) at *protein* level for MStern blotting-specific proteins (blue trace), FASP-specific proteins (yellow trace), shared proteins (green trace) and theoretical distribution of the entire human proteome (dashed grey trace). **b)** Comparison of physico-chemical property changes at *peptide* level: Molecular Weight (top), isoelectric pH (middle) and GRAVY score (bottom) for MStern blotting-specific peptides (blue trace), FASP-specific peptides (yellow trace), shared peptides (green trace) and theoretical distribution upon tryptically digesting the entire human proteome assuming no missed cleavages (0 MC; dashed dark grey trace) or 2 missed cleavages (2 MC; dashed light grey trace).

2.3.5 Protein quantification

Since this degree of missed cleavages will affect the quantification of individual peptides that are affected by being missed cleaved, the effect on the quantification of proteins was investigated, which normally uses the combined information from numerous peptides. To this end, two technical repeats of the HeLa cell lysates, neat urine and neat CSF digested using the MStern blotting and the FASP process (*cf.* **Figure 2.5**) were examined more closely.

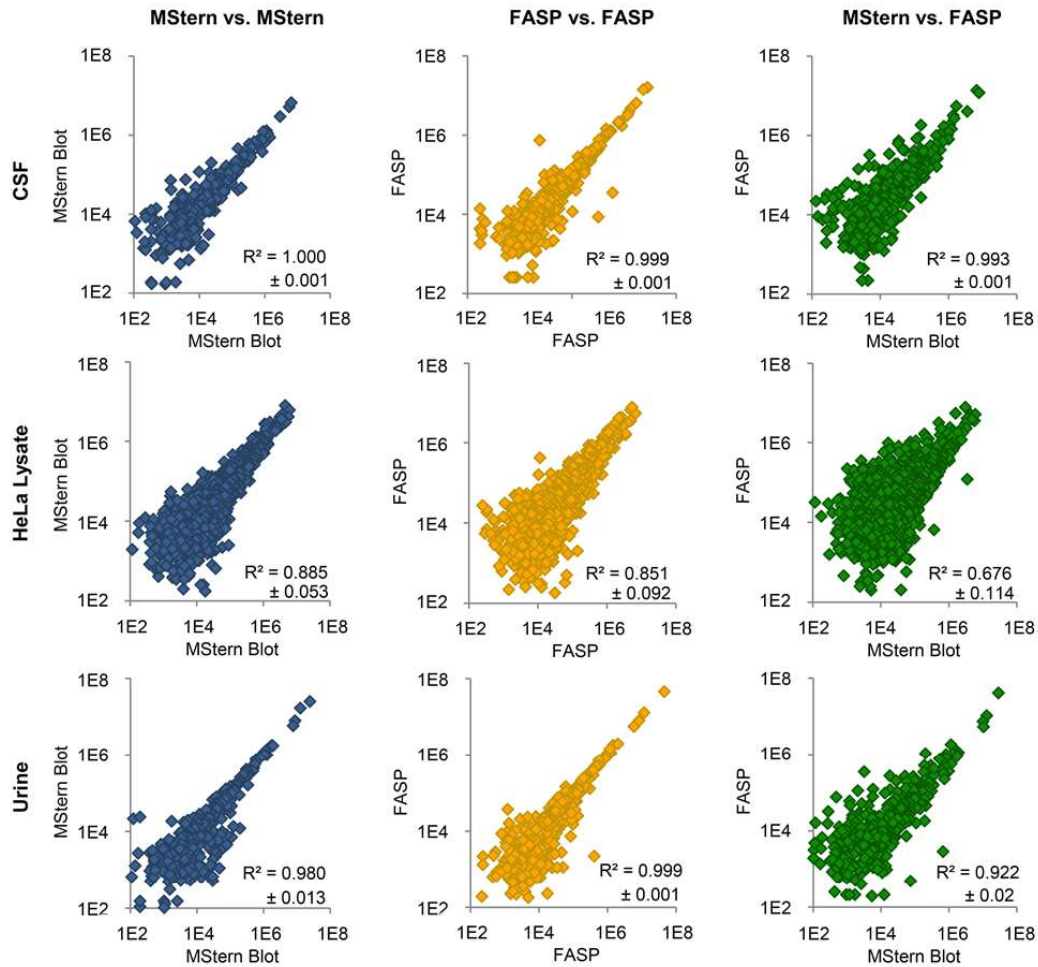


Figure 2.5: Correlation of FASP- and MStern Blotting-based Protein Quantifications. Correlation of the ProteinPilot-derived signal intensities of the proteins identified in CSF, HeLa lysate and urine (see Figure 2): MStern blot vs. MStern blot (left), FASP vs. FASP (middle) and MStern blot vs. FASP (right).

Next, the peptide ion signal intensity for each protein was extracted, prior to correlating the intensities for MStern blotting vs. MStern blotting (blue), for FASP vs. FASP (yellow) and for FASP vs. MStern blotting (green). The correlations for MStern vs. MStern and FASP vs. FASP were very tight with R²-values ranged from 0.85 to 1.0. The lower correlation for the HeLa lysate had to be expected given the complex nature of the samples, which results in undersampling, highlighting the negative effect of the stochastic nature of unbiased data-dependent acquisition routines on protein quantification, which is particularly limiting in the case of low abundant proteins. However, this limitation is independent of the sample processing, but can probably be improved when using non-stochastic data-independent acquisition routines for example.

The correlation of MStern vs. FASP showed a slightly broadened scatter with R²-values ranging from 0.92 to 0.99 for urine and CSF and 0.67 for the HeLa cell lysate. Such slight reduction in correlation is expected when comparing two independent sample processing method; nevertheless, the good to excellent correlations of the MStern vs. FASP-based quantification clearly shows that the increase in missed cleavage sides as observed for MStern blot-based processing still provides solid quantitative information comparable to and compatible with FASP-based processing.

2.3.6 Ovarian Cyst Biomarker Study

Using this newly developed MStern blotting strategy, 89 individual urine samples were processed (150 μ l of each), which were collected from pediatric patients of the Emergency Department at Boston Children's Hospital seen for abdominal pain with the aim to identify urinary markers for the different causes of abdominal pain. In this particular study, we were looking for urinary markers for ovarian cysts. While ovarian cysts in children and adolescent are common, normally asymptomatic, and easily detected by ultrasound, it is crucial to differentiate ovarian cysts from other conditions that require surgical intervention in a timely fashion, such as appendicitis or ovarian torsion. Thus, other means than slow imaging to quickly differentiate non-surgical conditions such as ovarian cysts from surgical conditions are highly desirable. To this end, neat urine samples, i.e. samples without any preprocessing and/or protein concentrations determination were processed using our MStern blotting strategy. The entire processing from the neat urine sample to the LC/MS ready sample took less than a workday (*cf.* **Figure 2.1**).

After analyzing these 89 samples with a 75 min gradient using a Q-Exactive mass spectrometer, the data were searched with MaxQuant¹¹⁷ and quantified using the iBAQ approach¹¹⁶. Before further analysis, 5 samples (3 ovarian cyst samples) were excluded in which less than 200 proteins were identified, probably due to low urinary protein concentrations. Of the remaining 84 urine samples, 10 were associated with ovarian cysts and 74 with alternative causes of abdominal pain. In total, 2,070 urinary proteins were identified in this sample set. The following statistical analysis (t-test) identified 11 proteins, which indicated statistical significance even after extremely conservative Bonferroni correction for multiple testing.

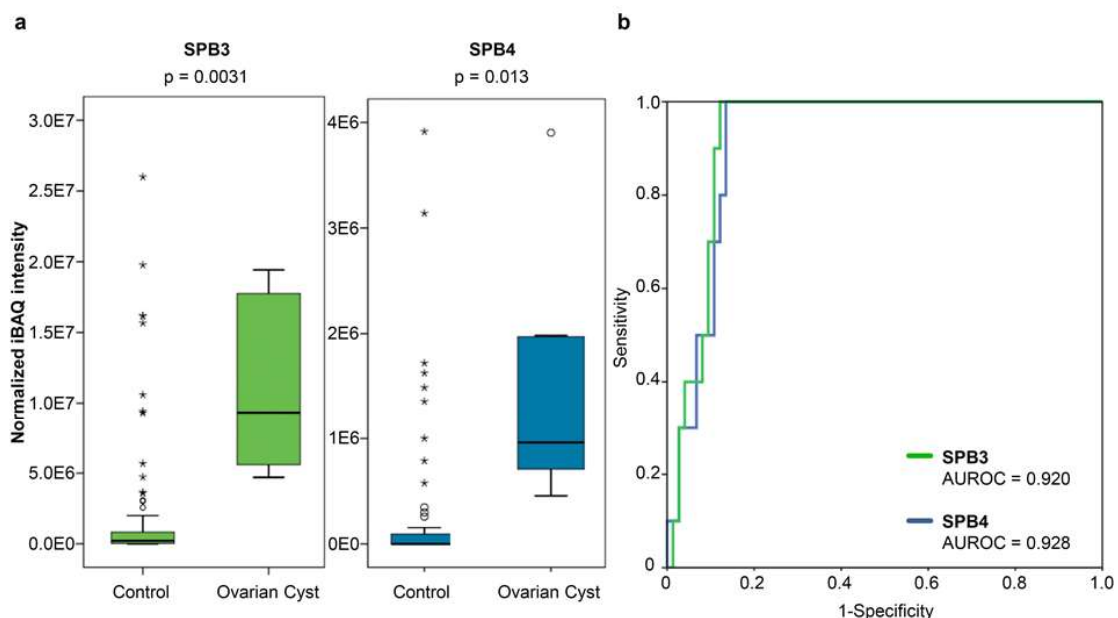


Figure 2.6: Statistical Analysis of Serpin B3 and 4 iBAQ intensities in urine. a) Box plot diagrams of the normalized iBAQ intensities of urinary serpin B3 and B4; listed p-values are Bonferroni corrected. b) Receiver operating characteristic (ROC) curves for urine serpin B3 and B4 based on 10 ovarian cyst cases and 74 abdominal pain controls.

Amongst these proteins that showed statistical significant abundance differences were two related serpins, namely serpin B3 and B4, also known as squamous cell carcinoma antigen 1 and 2 (cf. **Figure 2.6a**). Although serpins, in particular serpin E1, have been associated with e.g. polycystic ovary syndrome ^{126, 127}, serpin B3 and 4 have thus far only been described in cystic teratomas of the ovary ¹²⁸. Further studies beyond the scope of this work will be necessary to understand the biological and biomedical implications of this finding. However, further analysis of the ROC curves and box plots (AUROC SPB3 = 0.920, SPB4 = 0.928; **Figure 2.6b**) might indicate a promising application of these proteins as a rule-out test of ovarian cysts in pediatric patients.

2.3.7 Digestion Optimization with Trifluoroethanol

As previously mentioned, the newly developed MStern Blot approach demonstrates enhanced numbers of missed cleaved peptides. A thorough analysis of the raw data, especially for CSF digests, revealed the acquisition MS² spectra of highly charged (> 3+ charges) precursors. The occurrence of the highly charged precursor ions was additionally accompanied by high-intensity elution peaks in the higher organic region (>15% ACN) of the applied chromatographic gradient. These adverse occurrences gave rise to an incomplete digestion, particularly in case of CSF digestions. The initial experiments involved prolonged incubation times and increased organic content (up to 40% ACN) during digestion, which however did not prove a proper digestion of CSF.

As already described by Dickhut *et al.* (2014) ¹²⁵, the prolonged duration of incubation is not necessarily beneficial for digestion efficiency, but rather the enhancement of organic content. Dickhut *et al.* introduced the usage of Trifluoroethanol (TFE) as additional organic content during digestion to enhance the efficiency. They demonstrated the usefulness of TFE as an additive during digestion to circumvent barred cleavages due to phosphorylation at to trypsin adjacent amino acid residues. Here, the impact of TFE was investigated as a digestion additive on the overall digestion performance.

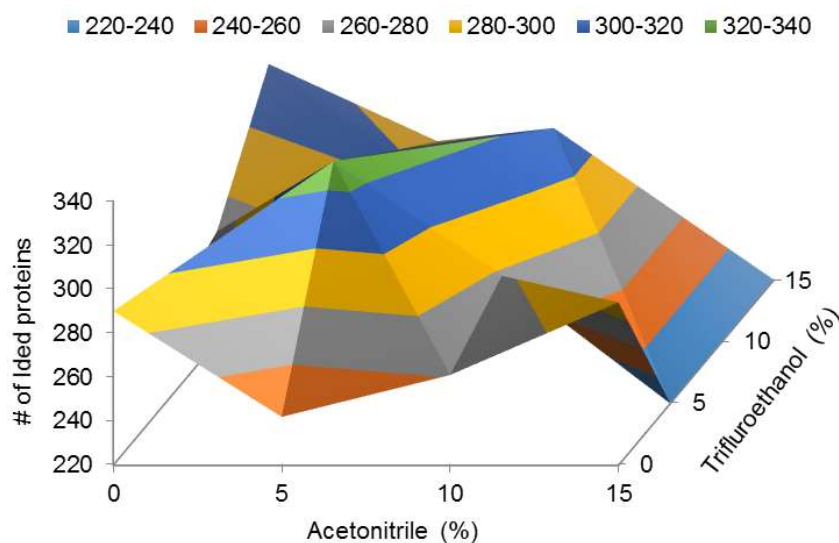


Figure 2.7: Impact of TFE on CSF protein identification. Different organic content compositions were tested during digestion: Acetonitrile (0 to 15%) with or without Trifluoroethanol (0 to 15%). Compositions exceeding a total organic content of 20% (e.g. 15% ACN with 5% TFE) have been omitted. The number of identified CSF proteins is illustrated.

Here, 12 different compositions of acetonitrile and trifluoroethanol were investigated as additive during digestion. Compositions that exceeded a total organic content of 20% (e.g. 10% ACN and 15% TFE) were omitted as the risk of PVDF membrane disintegration increases proportional to the amount of organic content the membrane is exposed to. As measurement of digestion efficiency, the number of identified proteins out of the identical and simultaneously processed CSF sample was used. Results for the diverse compositions are listed in **Table 2.1**.

Table 2.1: Number of identified proteins per applied organic content composition

Org. Content	0% TFE	5% TFE	10% TFE	15% TFE
0% ACN	290 ± 19	242 ± 38	261 ± 16	294 ± 22
5% ACN	258 ± 38	330 ± 36	278 ± 55	n/a
10% ACN	270 ± 57	306 ± 49	317 ± 25	n/a
15% ACN	318 ± 29	283 ± 67	n/a	n/a

As demonstrated in **Figure 2.7**, the best digestion conditions are achieved by supplementing the original digestion buffer (5% ACN) with 5% TFE. Hence, a total organic content of 10% (i.e. 5% ACN with 5% TFE) in the digestion buffer benefits the digestion of CSF within two hours, so that in average 330 ± 36 proteins were identified. Nonetheless, further applicable compositions have been observed. The digestion with either solely 15% ACN or 10% ACN with 10% TFE resulted in 318 ± 29 or 317 ± 25 identified proteins, respectively. These altered digestion conditions have also been tested for urine as sample. However, neither an improvement nor a debasement in protein identifications has been observed. Hence, it is postulated that these altered conditions are especially useful and should be applied while tryptically digesting CSF and/or plasma/serum with MStern Blot.

2.4 Conclusion and Perspective

Exploiting the high protein binding capacity of hydrophobic PVDF, which is also commercially available in the form of 96-well filtration plates, a 96-well plate-based sample processing method was derived, which allows for the complete processing of multiples of 96 samples in a workday or less. The major time advantages compared to e.g. FASP-based protocols are the fast liquid transfers and the omission of the need for desalting digests prior to loading onto LC/MS system. The former is the result of the pores which are 100 times larger than ultrafiltration membranes with appropriate molecular weight cut-offs. The latter was enabled by the efficient elution with organic solvents instead of high salt concentrations. This accelerated sample processing allows for the generation of LC/MS-ready peptide samples, starting from ~150 µl of neat urine, i.e. ~15 µg of protein, in a workday or less. Although no more than 15 µg of protein should be processed in a single well, this amount is easily sufficient for modern LC/MS systems, onto which less than 1 µg is normally injected for each run.

The direct comparison with FASP-based processing shows that our MStern blot processing results in at least as many proteins as FASP, with an overlap of identified proteins in the 65-75% range, although both methods show some process-specific biases. Although MStern blot results in an increase in missed cleaved peptides, which will alter the quantification of peptides affected by the missed cleavages, it can clearly be shown that the quantification of proteins, which is a composite value based on numerous peptides, is not affected by this increase in missed cleaved peptides. Another major advantage of the MStern blot method is the easy compatibility with liquid handling systems as liquid transfer is achieved using a vacuum manifold instead of a centrifuge which is necessary for FASP-based or other sample processing protocols for example¹²⁹.

An example application of the MStern blotting strategy is presented, in which 89 urine samples were processed within a single workday. The subsequent analysis of these urine samples identified two serpins as potential biomarker candidates for ovarian cysts based on their highly significant differential abundance in urine from ovarian cyst patients vs. abdominal pain controls.

In summary, MStern blotting is an ideal method to process dilute samples such as neat urine for downstream proteomic analysis, which lends itself to easy automation. Even though a particularly advantageous application for diluting samples such as urine has been noted, MStern is applicable to a wide range of samples without sacrificing analytical depth or quantitative nature of the data.

2.5 Acknowledgments and Contributions

The “MStern Blotting” technique was published in “Molecular Cellular Proteomics” as “Technological Innovations and Resources” article in October 2015 and has the following digital object identifier (doi):10.1074/mcp.O115.049650. Since the 1st of October 2016, the article has been available as open access resource article at “PubMed Central” with the PubMed Central ID (PMCID): PMC4597154.

I want to thank Dr. Judith Steen and Dr. Hanno Steen for insightful discussions and critically reading the manuscript. Furthermore, I want to thank Saima Ahmed, M.Sc., for the support while carrying out the aforementioned experiments. I want to thank Nerea Cuevas Polo for the help during cultivating the HeLa cells in cell culture. A big thank you also goes to Dr. Jan Muntel for the opportunity to validate this technique on his urinary biomarker study. Last but not least, I want to thank Dr. Richard Bachur and Dr. Alex Kentsis for the establishment of the urinary sample cohort and the collection of urine samples in the Emergency Medicine Department at Boston Children’s Hospital (Boston, MA, USA).

A special thank you goes to the “Technology and Innovation Development Office” at Boston Children’s Hospital for filing a patent application at the “United States Patent and Trademark Office” (USPTO). The patent status as of today is entitled as “Patent pending” with the international application number: WO 2016/112253 A1.

My personal contributions to this project account for the compilation of the initial idea, carrying out the experiments, analyzing the samples, data analysis, figure generation, writing the manuscript and scientific support during filing the patent application.

3. Urinary Biomarker Discovery for mTBI

3.1 Introduction

Each year, in the United States alone, traumatic brain injuries (TBI) account for approximately 2.2 million emergency department (ED) visits, more than 250,000 hospitalizations and 50,000 deaths¹³⁰. While ED visits for TBI continue to rise each year¹³¹, these numbers do not completely reflect the disease burden of mild TBI (mTBI) or concussion, which often remains undiagnosed or unreported¹³²⁻¹³⁴. Indeed, The Centers for Disease Control and Prevention (CDC) have described mTBI as reaching epidemic proportions, with an estimated 3.6 million sports-related traumatic brain injuries that occur each year in the United States^{135, 136}. Initial concussion management assessments can be critical, especially for school-aged children and athletes, as premature “return-to-play” decisions may increase the risk of worsened concussion symptoms or a second impact syndrome¹³⁶. However, post injury management is based largely on subjective symptom reporting, which can be challenging for clinicians to interpret and act upon, leading to wide practice variation in recommendations for activity cessation and specialized follow up. In fact, both

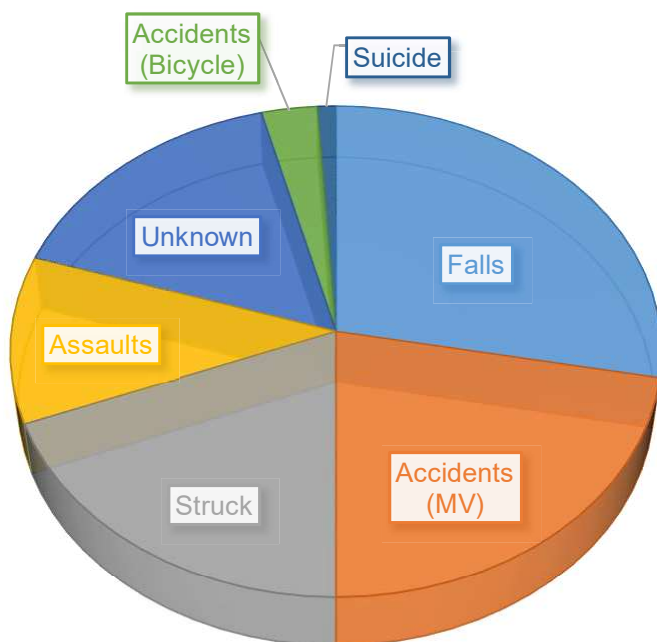


Figure 3.1: Causes of Traumatic Brain Injury. More than a quarter (28%) of all TBI cases are caused by falls. TBIs though accidents with motor vehicles (MV) and strucks demonstrate an incident of ~20%. Assaults account for about 11%, bicycle accidents for about 3% and suicide attempts for about 1% of observed TBIs. Approximately 16% remain unknown. Figure after <http://wikimedia.org>

pediatric and emergency care providers, who are most likely to diagnose and manage pediatric patients with sports-related concussion, have cited inadequate knowledge as a barrier to diagnosing and treating children with concussion⁸². A recent study in a pediatric tertiary care center demonstrated that nearly 60% of children who met diagnostic criteria for concussion were not diagnosed as such⁸⁴. Objective measures of injury and recovery after pediatric mTBI are sorely needed.

Potential diagnostic and prognostic biomarker candidates for acute mild traumatic brain injury have already been reported in prior studies by investigating different body fluids¹³⁷⁻¹⁴¹. Body fluids that have been investigated as source of biomarker for

concussion include cerebrospinal fluid (CSF), blood/plasma/serum, and saliva^{138, 142}. Although it has been argued that only CSF and/or blood are acceptable sources of biomarkers for mTBI, since too many barriers have to be crossed to make other fluids clinically viable¹³⁸, it remains unclear why crossing the blood-brain-barrier is clinically acceptable, but crossing the glomerular barrier in the kidney, for example, is not.

Urine is the one product of ultrafiltration of plasma in the kidney. Under normal physiological conditions large volume of plasma (350-400 ml/100g tissue/min) flows into the kidney, generating large amounts of ultra-filtrate (150/180 l/day) while urine output numbers only 1.5 l/day¹⁴³⁻¹⁴⁵. In short, urine as source of biomarkers for mTBI has the advantage of being non-invasively and readily obtainable in large volumes. Thus, any urine-based test is easily used by a PCP, a walk-in clinic, a primary health center, or even at home, in the sports facilities or on the battlefield. The effectiveness of urine tests in clinical care can be seen in the urine tests for drugs of abuse, pregnancy and glucose that have become standards of care in EDs throughout the United States.

3.2 Experimental Procedures

3.2.1 Urine Sample Collection

Urine samples were collected from consenting patients presenting to the ED at Boston Children's Hospital within 24 hours after concussive injury. After collection, the sample were directly transferred to the research facility by the associated research coordinators. Upon receipt the samples were aliquoted (depending on sample volume: 10x 2 ml; 1x 15 ml; 1x 25 ml and residual urine into 50 ml aliquots) and then frozen at -80°C. This study was reviewed and approved by the Institutional Review Board (IRB) at Boston Children's Hospital.

3.2.2 MStern Blot

Appropriate urine sample (e.g. 2 ml aliquote) was thawed and the neat urine (150 µl, i.e. ~15 µg of protein) was added to a mixture of 150 µg urea and 30 µl dithiothreitol (DTT) (100 mM in 1 M Tris/HCl pH 8.5). The resulting solution was incubated for 20 min at 27°C and 1100 rpm in a thermo mixer. Reduced cysteine side chains were alkylated with 50 mM iodoacetamide (IAA; final concentration) and incubation for 20 min in the dark at 27°C and 750 rpm.

The hydrophobic PVDF membrane in a 96-well plate format (MSIPS4510, Millipore) was pre-wetted with 150 µl of 70% ethanol and equilibrated with 300 µl urea supernatant (~8.3 M urea). These and all subsequent liquid transfers were carried out using a fitted 96-well micro-plate vacuum manifold (MAVM0960R, Millipore).

Each sample was transferred three times through the PVDF membrane by applying vacuum. After protein adsorption onto the membrane, it was washed twice with 50 mM ABC. Protein digestion was performed with sequencing grade trypsin (V5111, Promega) at a nominal enzyme to substrate ratio of 1:15. To this end, 100 µl digestion buffer (5% acetonitrile (ACN; v/v), 50 mM ABC and trypsin) were added to each well.

After incubation for 2 hours at 37°C in a humidified incubator, the remaining digestion buffer was evacuated. Resulting peptides were eluted twice with 150 µl of 40% ACN (v/v)/0.1% (v/v) formic acid (FA) each. Upon pooling, the peptide solutions were dried in a vacuum concentrator. Lyophilized samples were stored at -20°C for further analysis.

Table 3.1: Vendor specific entitlements for acquisition modes.

Acquisition Mode	Vendor	Entitlement
Data-dependent acquisition (DDA)	Thermo Fisher	Data dependent MS/MS fragmentation (ddMS ²)
	Sciex	Information dependent acquisition (IDA)
Data-independent acquisition (DIA)	Thermo Fisher	Data independent acquisition (DIA)

	Sciex	Sequential Windowed Acquisition of All Theoretical MS (SWATH)
--	-------	---

3.2.3 LC-MS/MS Analysis

Samples were dedicated for quantitative mass spectrometric data analysis. Here, “Data-Independent Acquisition” (cf. **Section 1.1.2.2**) was selected as the quantitative mass spectrometric method of choice. As described before, the quantification via DIA requires a comprehensive ion library via prior acquired DDA data files. DDA as well as DIA modes are entitled differently by each vendor and are further explained in **Table 3.1**.

3.2.3.1 Information Dependent Acquisition (IDA) for Ion Library Generation

Peptides were reconstituted in loading buffer (5% ACN (v/v), 5% FA (v/v)). LC-MS/MS analysis was performed on a microfluidic chip system (EK425) coupled to a TripleToF 5600+ mass spectrometer (both Sciex). Tryptic digests (~1 µg) were loaded onto a trap column (ReproSil-Pur C18-AQ, 200 µm x 0.5 mm, 3 µm) and subsequently separated on a ReproSil-Pur C18-AQ analytical column chip (75 µm x 15 cm, 3 µm) at a flow rate of 300 nl/min. A linear gradient from 95% to 65% buffer A (0.2% FA in HPLC water; buffer B: 0.2% FA in ACN) within 60 min was applied. Samples were ionized applying 2.3 kV to the spray emitter. Analysis was carried out in a data-dependent mode. Survey MS1 scans were acquired for 200 ms. The quadrupole resolution was set to ‘UNIT’ for MS2 experiments, which were acquired for 50 ms in ‘high sensitivity’ mode. The following switch criteria were used: charge: 2+ to 4+; minimum intensity: 100 counts per second (cps). Up to 35 ions were selected for fragmentation after each survey scan. Dynamic exclusion was set to 17 s.

3.2.3.2 Data-dependent Acquisition for the Biomarker Qualification Ion Library

Peptides were reconstituted in loading buffer (5% ACN (v/v); 5% FA (v/v)) and pooled based on their group affiliation. Samples originating from concussive samples were pooled into 6 LC-MS ready samples. Four control peptide mixtures were pooled into three samples to be analyzed and four intracranial bleed (IBD) samples resulting in three LC-MS ready samples. Pooled tryptic digests (~1.2 µg) were directly injected onto a PicoChip (NewObjective, ReproSil-Pur C18 3 µm, 75 µm x 105 mm) and peptides were consequently separated by a linear gradient from 98 % buffer A (0.2 % FA in HPLC water; buffer B: 0.2 % FA in ACN) to 70 % within 55 min and from 70 % to 65 % buffer A within 5 min. Analytes were ionized and sprayed into the mass spectrometer (Q Exactive, Thermo Fisher) by applying 2.2 kV to the spray emitter. The mass spectrometer was operated in data-dependent TOP10 mode with the following settings: mass range 375 – 1400 Th; resolution for MS1 scan 70 000 @ 200 Th; lock mass: 445.120025 Th; resolution for MS2 scan 17 500 @ 200 Th; max. injection time (IT)120 ms; isolation width 1.6 Th; Normalized Collision Energy (NCE) 27; underfill ratio 2.0 %; charge state exclusion: unassigned, 1, >6; dynamic exclusion 30 s.

3.2.3.3 Sequential Windowed Acquisition of All Theoretical MS (SWATH)

Peptides were reconstituted in loading buffer (5% ACN (v/v), 5% FA (v/v)). Peptide content for each prepared sample was measured via light adsorption at $\lambda = 280$ nm (Thermo Fisher NanoDrop 2000). LC-MS/MS analysis was performed on a microfluidic chip system (EK425) coupled to a TripleTOF 5600+ mass spectrometer (both Sciex). In short, Tryptic digests, normalized for peptide content (~ 1.2 μg), were loaded onto a trap column (ReproSil-Pur C18-AQ, 200 μm x 0.5 mm, 3 μm) and subsequently separated on a ReproSil-Pur C18-AQ analytical column chip (75 μm x 15 cm, 3 μm) at a flow rate of 300 nL/min. A linear gradient from 95% to 65% buffer A (0.2% FA in HPLC water; buffer B: 0.2% FA in ACN) within 60 min was applied. Samples were ionized applying 2.3 kV to the spray emitter. Analysis was carried out in data-independent mode (SWATH). Each DIA cycle contains one full MS scan and 24 consecutive DIA scans covering a mass range of 400 to 1000 Th covering 97% of peptides¹⁴⁶ in a urinary sample. All fragmentation windows featured a fixed window size of 25 Da with an overlap of 1 Da with the previous window. The MS¹ scan was setup to scan all masses between 400 m/z and 1250 m/z within 250 ms. Consecutive MS² spectra, containing fragment ions from all precursors falling into the dedicated precursor mass range (25 Da), were acquired along a 250 m/z to 1500 m/z mass range for a duration of 125 ms in “High Sensitivity” mode. Resolution for all MS² spectra were set to “UNIT”. The collision energy applied for each fragmentation window was calculated on-the-fly by an implemented algorithm designed by Sciex. Collision energy algorithm was applied with default values specified by Sciex.

3.2.3.4 DIA Analysis for Biomarker Qualification

Peptides were reconstituted in loading buffer (5% ACN (v/v); 5% FA (v/v)) and concentrations normalized by spectroscopically (NanoDrop 2000) determined peptide content. Required normalization peptides (HRM Calibration Kit; Biognosys, Switzerland) were spiked in a 1:30 ratio. Tryptic peptide mixtures (~ 1.2 μg) were directly injected onto a PicoChip (NewObjective, ReproSil-Pur C18 3 μm , 75 μm x 105 mm) and peptides were consequently separated by a linear gradient from 98 % buffer A (0.2 % FA in HPLC water; buffer B: 0.2 % FA in ACN) to 70 % within 55 min and from 70 % to 65 % buffer A within 5 min. Analytes were ionized and sprayed into the mass spectrometer (Q Exactive, Thermo Fisher) by applying 2.2 kV to the spray emitter. Each DIA cycle contained one full scan (MS¹) and 15 consecutive parallel reaction monitoring (PRM) scans covering a mass range of 400 – 1180 Th. The full scan resolution was set to 35,000 @ 200 Th; AGC target $-3\text{e}6$, maximal IT 120 ms; mass range 400 – 1,200 Th; followed by DIA scans with resolution 35,000 @ 200 Th; variable window size was applied used; NCE 27; AGC target value $1\text{e}6$, max. IT 120 ms.

Table 3.2: Variable extraction windows used in DIA experiments.

Window No.	Window size
1	52 Da @ 405 m/z
2	42 Da @ 450 m/z

3 to 8	32 Da @ 485, 515, 545, 575, 605 and 635 m/z
9 and 10	37 Da @ 687.5 and 702.5 m/z
11	42 Da @ 740 m/z
12	47 Da @ 782.5 m/z
13	57 Da @ 832.5 m/z
14	82 Da @ 900 m/z
15	162 Da @ 1020 m/z

3.2.4 Data Analysis

3.2.4.1 DIA Analysis for Biomarker Discovery

The generated DDA files (*.wiff) (cf. **3.2.3.1 Information Dependent Acquisition for Biomarker discovery**) were loaded collectively into ProteinPilot™ (version 4.5.1, revision 2768; Paragon™ Algorithm 4.5.1.0, 2765; Sciex) and searched using the human UniProtKB database (Homo sapiens, ~20,300 sequences, version 06-2014) as reference. The ‘thorough’ search mode was used with the following settings: Cys Alkylation – Iodoacetamide; Digestion – Trypsin; Instrument – TripleTOF5600; Special Factors – Urea Denaturation; Species – None; FDR analysis requested. Commonly occurring laboratory contamination protein sequences (cRAP, version 2012.01.01) were added to the database. In further steps, the generated database search result file (*.group) was used as ion library for the analysis of DIA files.

The quantitative analysis of DIA files was performed via “MS/MS(ALL) with SWATH™ Acquisition MicroApp” (Sciex, v.1.0) plugin in PeakView™ (Sciex, v.1.2). In short, the according DDA search result file (*.group) was loaded into the PeakView™ MicroApp as spectral library. As the maximal number of proteins to be imported for analysis, the number of identified proteins at 1% global FDR out of the automatically generated FDR analysis result file (byproduct of the ProteinPilot™ search algorithm), was specified. For a more stringent analysis, shared peptides were excluded and based on the experiment design, “unlabeled” proteins were specified. After spectral library generation, the corresponding DIA files were loaded into the PeakView™ MicroApp. The DIA files were processed with default settings, the “Peptide Confidence Threshold” was adjusted according to the findings in the FDR analysis. The protein quantitation result file (*.mrkvw) was opened in MarkerView™ (v.1.2.1, Sciex) and values copied for downstream analysis.

Statistical analysis was performed either in R, SPSS v21 (IBM) or EXCEL with a “Real Statistics Resource Pack” plugin. The significance of protein abundance changes was calculated using the nonparametric Mann–Whitney u-test with Benjamini-Hochberg multiple testing correction. Data manipulation as imputation was avoided. Before statistical analysis, the dataset was sample sum average normalized to avoid an impact of LC-MS run specific intensity variations on the calculated p-values.

3.2.4.2 DIA Analysis for Biomarker Qualification

Urinary samples for the biomarker qualification (extended sample cohort) was analyzed on a Thermo Fisher QExactive. Generated Thermo Fisher generic data files (*.raw) were analyzed differently, as PeakView™ and the corresponding plugin (both Sciex) do not support the analysis of mass spectrometric data files originating from other vendors.

For ion library generation, the corresponding ten pooled DDA analysis files (*.raw) were loaded into MaxQuant¹¹⁷ (v.1.5.2.8) and consequently searched against a reference protein sequence database, here the human UniProtKB database (Homo sapiens, ~20,300 sequences, version 06-2014) was used with commonly occurring laboratory contamination protein sequences (cRAP, version 2012.01.01) added. Variable Modifications were set to: Acetyl(Protein N-Term), Oxidation(M), Deamidation(NQ) and Phospho(STY). For fixed modifications only Carbamidomethyl(C) was used. The digestion mode was set to “Semispecific” and the digesting enzyme to ‘Trypsin/P’. Otherwise default settings (by MaxQuant) were used for the analysis: trypsin with up to three missed cleavages; mass tolerance for the first search: 20 ppm; main search: 4.5 ppm.

DIA files (*.raw), which belong to the analysis of the larger cohort for biomarker qualification, were processed and analyzed in Spectronaut¹⁴⁷ (v.9.0.1 (Collins); Biognosys). At first, the search results from the aforementioned MaxQuant search of the corresponding DDA acquisition were loaded into Spectronaut to generate an ion library, which in turn was saved within the program’s data-structure. The data ion library import was performed via the default settings, proposed by Spectronaut. Nonetheless, peptide and protein grouping was kept on the results by the Andromeda search engine (MaxQuant). Hence, the “Protein Inference” option was deactivated for ion library generation. After generating the ion library, corresponding DIA files were loaded into Spectronaut and the aforementioned ion library was specified as a reference identification database. The following custom settings were applied in Spectronaut: Calibration Mode: Forced; Protein Quantity: Sum Precursor Quantity; Data Filtering: Either QValue or QValue Sparse. All results were filtered by a Q value of 0.01 (equals a FDR of 1% on peptide level). The output report was adjusted to contain all quantification-related values. All other settings were set to default. For further analysis, the peptide-based result file was chosen.

Statistical analysis was performed either in R, SPSS v21 (IBM) or EXCEL with “Real Statistics Resource Pack” plugin. Significance of protein abundance changes was calculated using the non-parametric Mann–Whitney u-test with Benjamini-Hochberg multiple testing correction. Data manipulation as imputation was avoided. Before analysis, the dataset was sample sum average normalized to avoid an impact of LC-MS run specific intensity variations on the calculated p-values.

3.3 Results and Discussion

3.3.1 Sample cohorts

In this study, it is evaluated whether or not potential biomarker targets can be found in the urine of children and adolescents between 11 and 22 years of age, presenting to the ED within 24 hours after concussive injury. Concussion subjects were defined as patients who sustained a blunt injury to the head resulting in either i) alteration in mental status (including loss of consciousness, disorientation, or amnesia) or ii) any of the following symptoms that started within 4 hours of the injury (and not present before the injury): headache, nausea, vomiting, dizziness/balance problems, fatigue, drowsiness, blurred vision, memory difficulty or difficulty concentrating¹⁴⁸. Subjects were excluded for the following criteria: intracranial hemorrhage noted on advanced imaging, Glasgow Coma Scale (GCS) less than 13, co-existing orthopedic fracture, co-existing intra-abdominal or intra-thoracic trauma, or spinal cord injury. Control patients were obtained from the same source population and included subjects with non-concussive minor head trauma (e.g. laceration or contusion, with no clinical signs of concussion) and subjects with isolated orthopedic injury (e.g. long bone fracture).

Biomarker Discovery Cohort

In the Biomarker Discovery study, a total of 22 patients was enrolled, belonging to three distinct patient groups: 8 concussion cases, 6 patients with minor head trauma (HI control), 6 patients with orthopedic trauma (FR controls) and 2 patients with intracranial bleeding (no specific subtype) as advanced/positive control. No particular focus was set onto the timeframe between initial injury and time of urine sampling. Additionally, concussive patients were not selectively chosen as sportive mTBI. In this initial study, all sorts of accidental mTBIs and TBIs were allowed for this study.

Biomarker Qualification Cohort

For the qualification of discovered biomarker candidates, the initial "Biomarker Discovery Cohort" was extended by a total of 27 samples. Here, 14 additional patients were enrolled who were suffering from concussion, 6 additional control samples (without any further stratification) and 7 supplemental patients for the positive control or i.e. intracranial bleed subgroup. These additional samples were provided with supplemental information in regard of "Time of Injury" and "Time of Urine Collection": samples originating from concussive patients were sampled in between $T_{\min} = 1 \text{ hr } 50 \text{ mins}$ and $T_{\max} = 13 \text{ hrs}$. Control samples featured similar sample times with $T_{\min} = 1 \text{ hr } 50 \text{ mins}$ and $T_{\max} = 9 \text{ hrs } 50 \text{ mins}$. The timeframe for samples originating from patients with intracranial bleed was comparably the longest with $T_{\min} = 4 \text{ hrs}$ and $T_{\max} = 20 \text{ hrs}$.

3.3.2 Biomarker Discovery

3.3.2.1 Ion library and Identifications

In total, 22 DDA files, acquired on the TripleTOF 5600 were searched with ProteinPilot™ which resulted in 1236 identified proteins at 1% global FDR. Furthermore, ProteinPilot™ identified and matched 239,523 acquired spectra (1% global FDR; sum across all submitted files) to a peptide sequence, present in the specified reference sequence database. This number of identified spectra can be summarized into roughly 25,000 distinct peptide species, whereas a different modification or modification combination is counted separately even by identical underlying peptide sequence. This identification information is consequently loaded into the ion library for matching SWATH MS² spectra peak patterns to those of already identified spectra.

3.3.2.2 SWATH-based Protein Biomarker Discovery

The DIA dataset features a total of 672 uniformly quantified proteins. The SWATH processing algorithm is programmed to only report proteins which could be quantified in all imported samples. Hence, the dataset does not feature any missing values. This particularly allows for the usage of the dataset as it is acquired and does not require any further manipulation for imputation or handling of missing values. However, this approach limits the extent of the screening depth as only rather high abundant proteins are reported as they are easily detected and identified in all samples

149

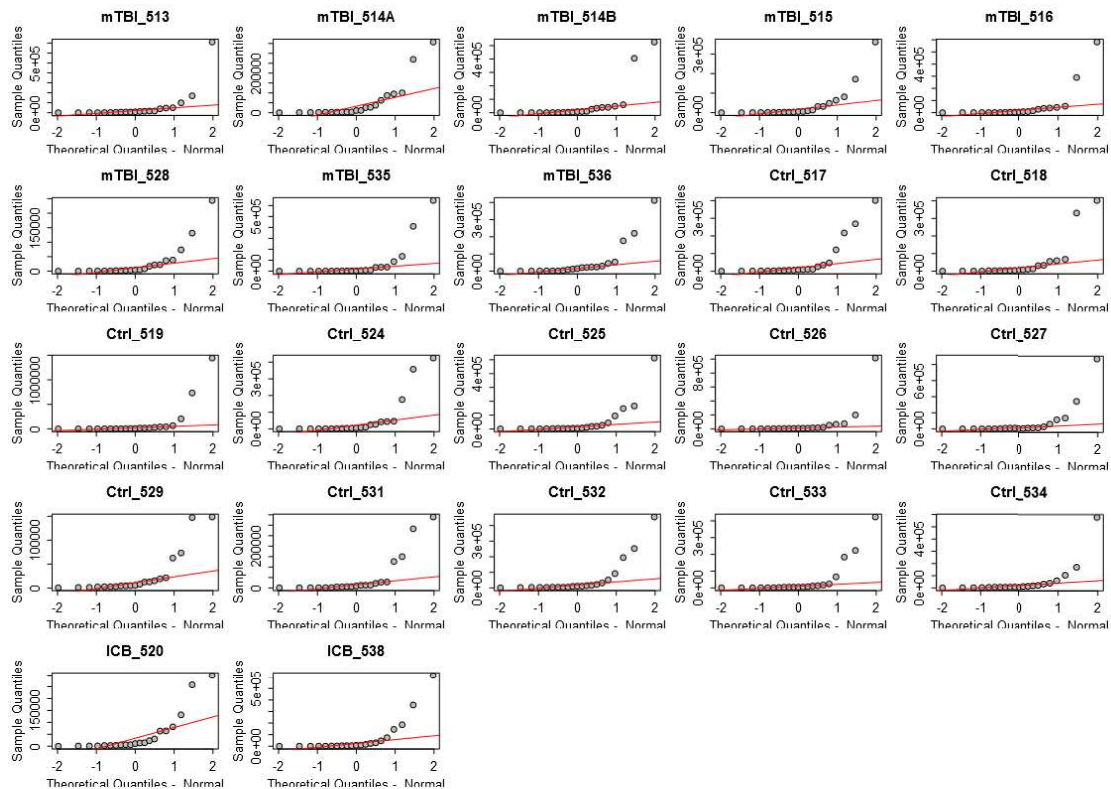


Figure 3.2: Q-Q-Plots for samples analyzed by SWATH. Sample quantiles are plotted against theoretical quantiles with an underlying normal distribution. Once data points form a diagonal, normal distribution for the sample has been proven. Any other graphical distribution provides evidence for non-parametric distribution. Sample abbreviations: mTBI – case samples for mTBI; Ctrl – control samples; ICB – positive control samples (intracranial bleeding).

After sum normalization (normalized to the average of the summed sample intensity), the data distribution was tested while illustrating the data with Q-Q-Plots (cf. **Figure 3.2**). Here, no evidence in any samples was found that gives rise for a normal data distribution. Hence, for any further analysis, a non-parametric test was chosen. As the first analysis step, the protein abundance changes between “Control” and “mTBI” group where tested for statistical significance by applying the non-parametric Mann-Whitney u test ¹⁵⁰. A total of 21 proteins were identified with a statistical significant (p-value < 0.05) protein abundance change (cf. **Table 3.3**).

Table 3.3: Observed proteins with significantly different protein abundance.-Accession are based on the UniProt terminology. Marked p-values refers to different types: * refers to “original” values; ϕ refers to Benjamini-Hochberg ¹⁵¹ multiple-testing corrected p-values.

Accession	Protein Name	p-value *	p-value ϕ
P08138	Tumor necrosis factor receptor superfamily member 16	0.00547855	0.9570165
P00738	Haptoglobin	0.00547855	0.9570165
Q06828	Fibromodulin	0.00871191	0.9570165
P20160	Azurocidin	0.01089635	0.9570165
Q9HD89	Resistin	0.01677046	0.9570165
P04180	Phosphatidylcholine-sterol acyltransferase	0.01677046	0.9570165
O43505	N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase	0.02526030	0.9570165
P80188	Neutrophil gelatinase-associated lipocalin	0.02526030	0.9570165
P05154	Plasma serine protease inhibitor	0.03075356	0.9570165
Q86T13	C-type lectin domain family 14 member A	0.03724254	0.9570165
P01611	Ig kappa chain V-I region Wes	0.03724254	0.9570165
O14773	Tripeptidyl-peptidase 1	0.03724254	0.9570165
P06889	Ig lambda chain V-IV region MOL	0.03724254	0.9570165
P26927	Hepatocyte growth factor-like protein	0.03724254	0.9570165
P04217	Alpha-1B-glycoprotein	0.04486227	0.9570165
Q99969	Retinoic acid receptor responder protein 2	0.04486227	0.9570165
P02753	Retinol-binding protein 4	0.04486227	0.9570165
O75882	Isoform 2 of Attractin	0.04486227	0.9570165
Q6P531	Gamma-glutamyltransferase 6	0.04486227	0.9570165
P48061	Stromal cell-derived factor 1	0.04486227	0.9570165
P54793	Arylsulfatase F	0.04486227	0.9570165

The whole list of proteins tested contained a total of 672 proteins. Corresponding p-values for the statistical significance of abundance change were consequently corrected for multiple testing by “Benjamini-Hochberg”¹⁵¹ multiple testing correction. Based on the multiple-testing corrected p-values, none of the previously identified proteins features a p-value < 0.05. Based on the fact that urine features high inter-personal variances¹⁵² and p-values itself are highly discussed in the context of biomarker discovery¹⁵³⁻¹⁵⁵, corrected p-values were disregarded for this biomarker discovery study.

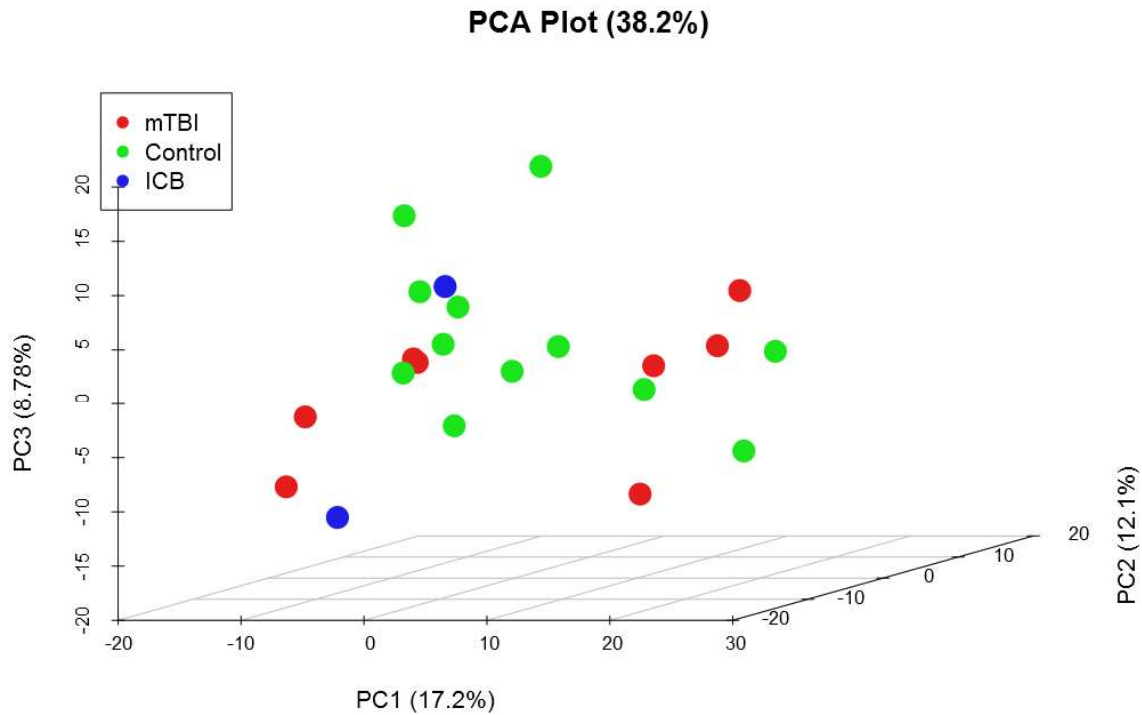


Figure 3.3: Principal Component Analysis for all quantified proteins. Principal Component 1 (17.2%), Principal Component 2 (12.1%) and Principal Component 3 (8.8%) account for a total of 38.2% of the overall variance in the dataset. No clear visual separation or clustering between “positive control” (ICB; blue), negative control (Control; green) and cases (mTBI; red) is apparent.

In the first candidate identification step, the whole list of quantified proteins was used for a “Principal Component Analysis” (PCA). Here, variations are combined into “components” with the same variation effect. Hence, the dimensions of possible variations are condensed into “principal components” (PC) and help to understand data variation effects and its related effect strength. In this context, PC 1 accounts for 17.2%, PC 2 for 12.1% and PC 3 for 8.78% of the whole dataset variation. Here, the top three components (PC 1, PC 2 and PC 3) collectively account for 38.2% of the whole dataset variation. Nonetheless, any linear combination (*cf.* **Figure 3.3** and **Appendix 8.2**) of each of those components (PCA plot) lead to no clear grouping, or i.e. group-based clustering. As a consequence, only proteins with a significantly different protein abundance (not multiple testing corrected; $p < 0.05$) were entered into the PCA, to evaluate any diagnosing tendency for these proteins. The consequent PCA plot for either PC1 and PC2 (*cf.* **Appendix 8.2**) or PC1 and PC3 (*cf.* **Figure 3.4**) demonstrates a clear group-based separation. This leads to the assumption that the proteins that have been found with a significant abundance change (non-corrected, $p < 0.05$) feature a certain diagnostic ability for concussion in urine.

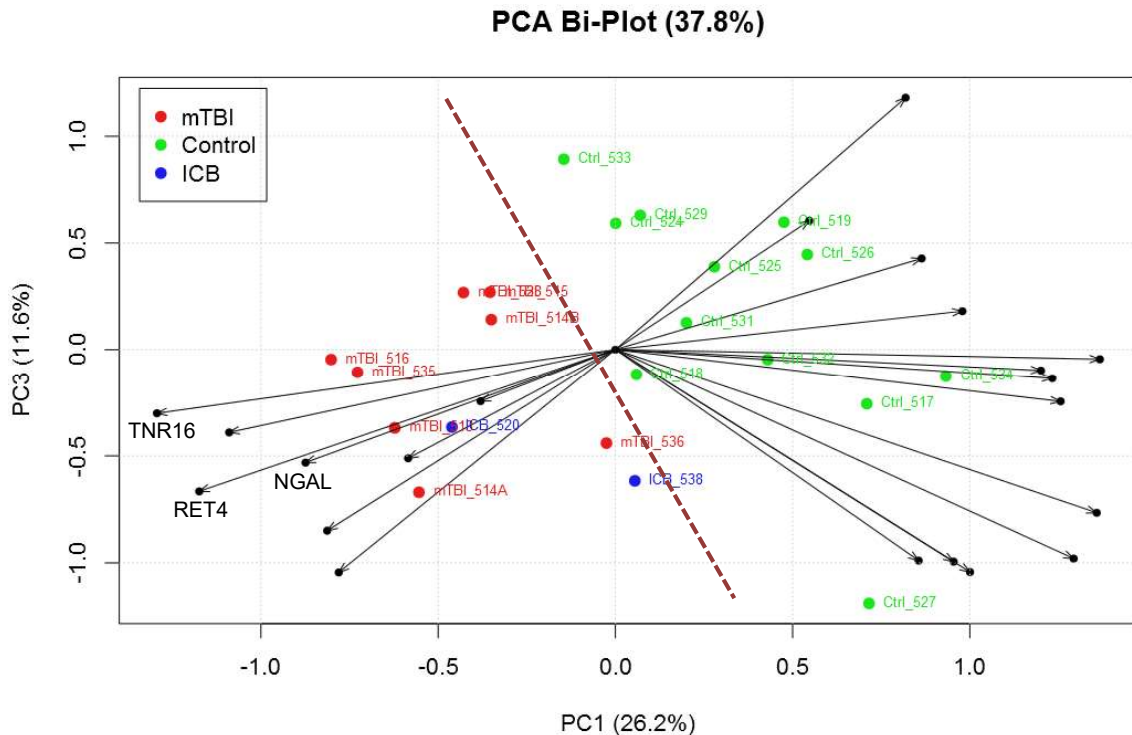


Figure 3.4: PCA Bi-Plot (PC1 and PC3) for significantly different abundant proteins. PCA Plot demonstrates PC1 (26.2% x-axis) and PC3 (11.6%; y-axis) of the PC analysis for significantly different abundance observation between Controls (green) and Cases (red). Principal components account for 37.8% of the total variance within the significant different abundant proteins. Black arrows demonstrate variance features and its effect strength and direction. Clear separation of cases (red and blue) and controls to one side of the plot (red dotted bar).

Amongst the identified proteins, three distinct proteins: i) Tumor Necrosis Factor Receptor Superfamily Member 16 (TNR16, p75^{NTR})¹⁵⁶⁻¹⁵⁸, ii) Retinol-binding protein 4 (RBP4, RET4)¹⁵⁹ and iii) Neutrophil gelatinase-associated lipocalin (NGAL, *Lcn2*)¹⁶⁰, have already been studied in the context of traumatic brain injury.

Previous studies^{156-158, 161}, have shown, that TNR16, or i.e. Nerve Growth Factor Receptor (NGFR, p75^{NTR}) signaling is involved directly after brain injury. After severe impact to the head, neurotrophins bind to NGFR located on neurons, signaling the cells to induce apoptosis and hence neuron degeneration and secondary brain damage. Additionally, p75^{NTR} is significantly overexpressed in neurons after brain injury and thus actually promoting neuron degeneration and secondary brain damage, which in turn promotes the usage of this particular protein as a prognostic tool of TBI.

Hergenroeder *et al.* (2008)¹⁵⁹ conducted a proteomic serum biomarker discovery study in brain-injured adults. Here, they established a connection between altered RET4 (RBP4) levels in serum and brain injury. Furthermore, they found that RET4 might also have a predicting ability of intracranial pressure (ICP) after severe traumatic brain injury, which in turn is a major contributor to death and disability.

Jeon *et al.* (2013)¹⁶⁰ used an *Lcn2* (NGAL) deficient mouse strain to investigate the role of NGAL after spared nerve injury (SNI). In this study, they found that *Lcn2* mRNA levels were elevated in the spinal cord and NGAL was localized in the spinal neuron after SNI. Based on these findings,

they concluded that *Lcn2*, i.e. NGAL, plays a crucial role in the spinal cord for the development of neuropathic pain following a peripheral nerve injury.

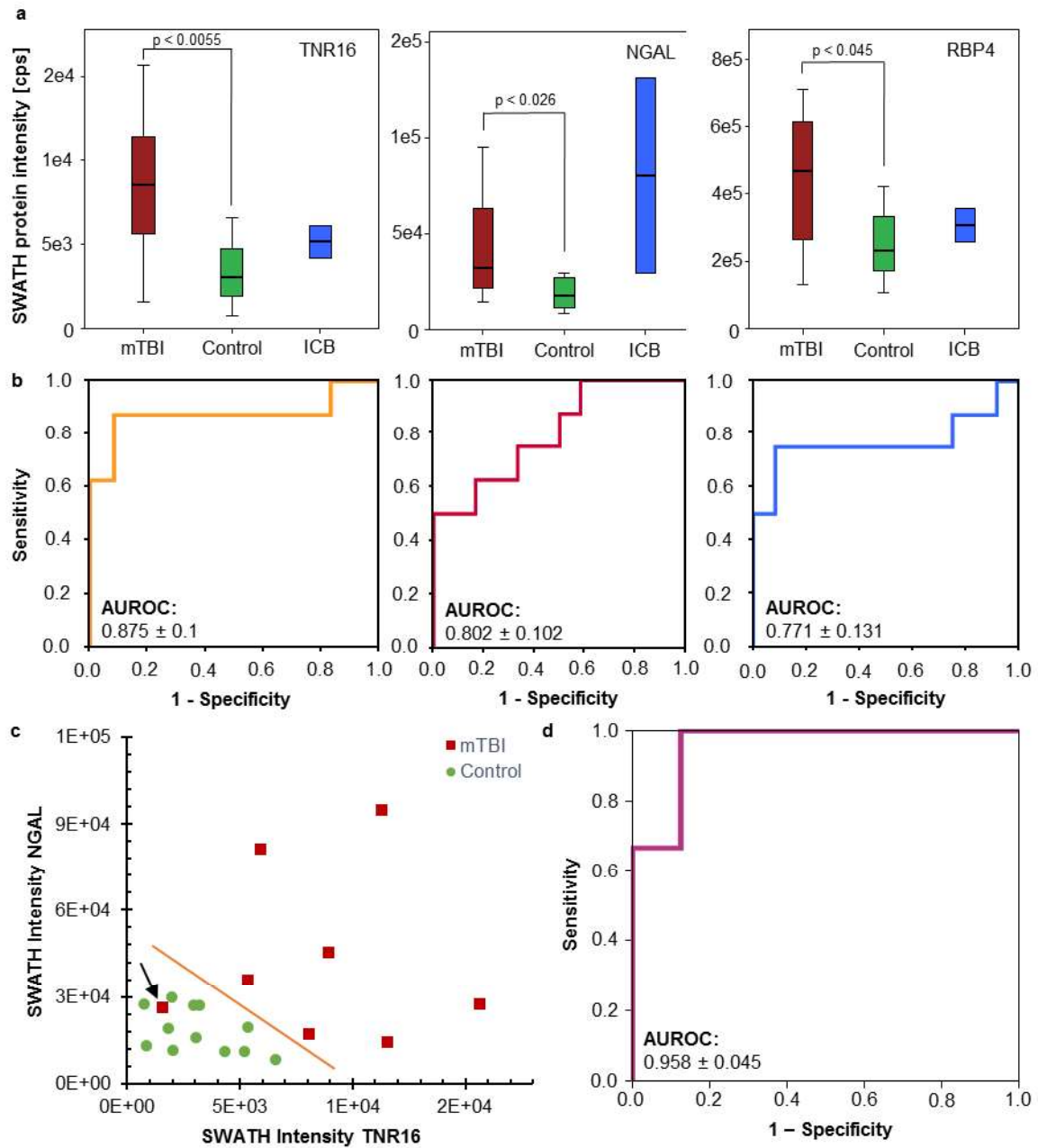


Figure 3.5: Evaluation of possible biomarker candidates. **a)** Boxplots are based on the observed protein abundance acquired by SWATH. The corresponding p-values are based on non-corrected Mann-Whitney u testing between mTBI and Control samples. Intracranial bleed samples are excluded in this consideration as the cohort is represented by an incomprehensive number of cases (2). **b)** ROC curves for diagnosing ability based on single proteins: TNR16 (yellow), NGAL (red), RBP4 (blue). **c)** Scatter plot of SWATH intensities for TNR16 (x-axis) and NGAL (y-axis). Distinct group-based clustering of samples belonging to either mTBI (red) or Control (green). Only a single sample (arrow) clusters with the opposite group. **d)** ROC analysis of the combined diagnosing ability of TNR16 and NGAL. Combined effect achieves an AUROC of 0.958 ± 0.045 . Almost perfect biomarker panel (AUROC = 1), alleviated due to sample clustering with opposing group.

As the three aforementioned proteins (TNR16, NGAL and RBP4) have already successfully been described in various contexts of mTBI and TBI, the applicability of these proteins as possible urinary biomarker candidates for concussion is further evaluated. The direct comparison between “Control” and “mTBI” group revealed a larger spread of mTBI cases, which in turn indicates that a clinical applicability might be compromised and a distinct group classification by a single protein cannot be achieved. For further evaluation for the use of single proteins for diagnosing mTBI in urine, the ROC curves were analyzed (*cf.* **Figure 3.5b**). For each protein alone, an area under the ROC curve (AUROC) of > 0.5 was achieved: 0.857 ± 0.103 , 0.723 ± 0.113 and 0.768 ± 0.133 for TNR16, NGAL and RBP4, respectively.

Besides evaluating the diagnosing efficiency of single proteins for diagnosing mTBI in urine, here, combinations of proteins and hence biomarker candidate panels were also evaluated for distinct diagnosing ability. As RBP4 (Retinol binding protein 4) has only been described briefly as being involved in mTBI¹⁵⁹ (increased ICP) as well as being only a possible indicator for a particular secondary mTBI condition, this protein has been excluded from being under consideration as a possible member of the diagnostic biomarker panel for mTBI.

Here, the more detailed analysis of the biomarker candidates identified in the current study that by combining the information of proteins, cases separate from the control samples. Particularly in the case of combining the protein abundance information of TNR16 and NGAL, all but one concussion case showed a strong separation from all controls in a 2D scatter plot (*cf.* **Figure 3.5c**). The resulting ROC analysis resulted in a strong area under the ROC curve of 0.958 (*cf.* **Figure 3.5d**). Ultimately, the combination of all three previously described proteins (TNR16, NGAL and RBP4) demonstrates perfect biomarker conditions with an AUROC of 1.

In summary, the above-described results underscore the notion that it might be possible to “objectively” diagnose concussion, based on concussion-specific changes in the urinary proteome composition. These biomarker candidates however, still need to be validated.

3.3.3 Biomarker Qualification

3.3.3.1 Ion library and protein identifications

In contrast to the previous “biomarker discovery” (*cf.* **Section 3.3.1**), the ion library for the biomarker qualification is based on an extended cohort. Besides the cohort extension, this data was also acquired on a different instrumentation (QExactive). Due especially to the use of different instrumentation and vendor specific data restriction, the data analysis software needs to be adjusted. In this regard, the previously acquired ion library (*cf.* **Section 3.3.2.1**) as well as the underlying DDA files are no longer applicable. Here, a new set of DDA analysis files had to be acquired, spiked with retention time calibration peptides (iRT Kit¹⁶², Biognosys, Switzerland). As mentioned previously, the ion library used for biomarker candidate discovery was established by a single DDA file for each sample. Hence, each sample was acquired once by a DDA acquisition routine and then in DIA mode. Consequently, a sample-by-sample acquisition within this cohort would require the acquisition of about 50 samples, which would then last multiple days. In order to minimize instrument usage as well as information redundancy, samples were pooled according

to their grouping. The pooled samples were consequently acquired on a QExactive (Thermo Fisher) mass spectrometer and analyzed by MaxQuant.

The import of the MaxQuant search results into Spectronaut was carried out without their built-in inference algorithm. It was observed, that by using this approach, duplicate search hits can occur

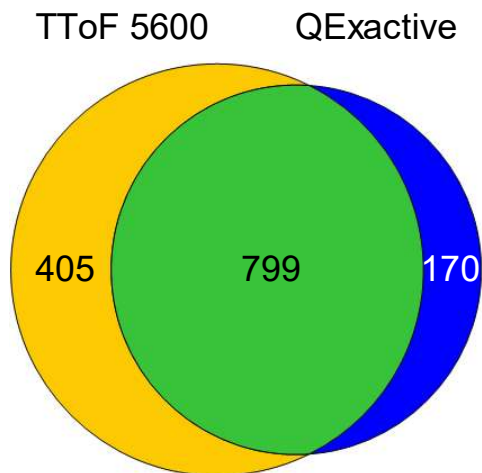


Figure 3.6: Venn diagram of used ion libraries.

Comparison of ion libraries used for protein quantification via DIA data. Given numbers represent the number of previously identified protein groups. After manual filtering of protein entries, the ion library used for SWATH quantification (yellow) features 1204 protein groups. In contrast, the QExactive (blue) ion library has 969 entries. A total of 799 proteins (58.2%; green) is shared between both libraries.

with different quantification values, leaving the opportunity for biasing the outcome. Here, the ion library featured 979 unfiltered protein groups, which account for a total of 1187 unique proteins. Based on the search results, 119 proteins or protein groups are only based on a single precursor ion. However, ions were filtered based on their assigned PSM FDR of 1% through MaxQuant. The assignment of this ion library to the workflow allows for tracing of 9432 precursors out of a total of 9462 identified precursors in MaxQuant.

The direct comparison of both ion libraries revealed that 58.2% of all protein identifications are shared between both libraries (*cf.* **Figure 3.6**). The initial ion library used for quantification of SWATH data originating from a TToF 5600 system had 405 (29,5%) additional entries, which is unique for this cohort. In contrast, the ion library used to quantify

DIA data originating from a Thermo Fisher QExactive mass spectrometer had 170 unique additional entries. Here, the coverage of the initial discovery cohort through the follow-up cohort at 58.2 % is sufficient to compare protein quantitation.

3.3.3.2 Protein quantifications

The protein quantification in the biomarker qualification study is based on the sum of the total MS² peak area for each assigned peptide. This method of quantification, however, allows missing values so that certain proteins feature “zero” abundance in certain samples. At first, protein quantities were sum normalized across the dataset to account for technically introduced quantity changes. Thereafter, the two most abundant proteins, namely Uromodulin (P07911) and Albumin (P02768) were removed from the protein list and hence from any further consideration. Additionally, any other specifically identified contaminants (CON) were also removed from the list and consequent statistical analysis. After protein filtering the list of quantified proteins consisted of 967 protein groups while each sample featured an average of 840 ± 64 quantified protein groups. This in turn characterizes each sample with an average of 15.8 % of missing values. Nonetheless, missing values represent the cases where the corresponding protein was either “not present” in this sample or the abundance was lower than the actual detection limit of the instrumentation. In further analyses, missing values were not further imputed and was kept as-is.

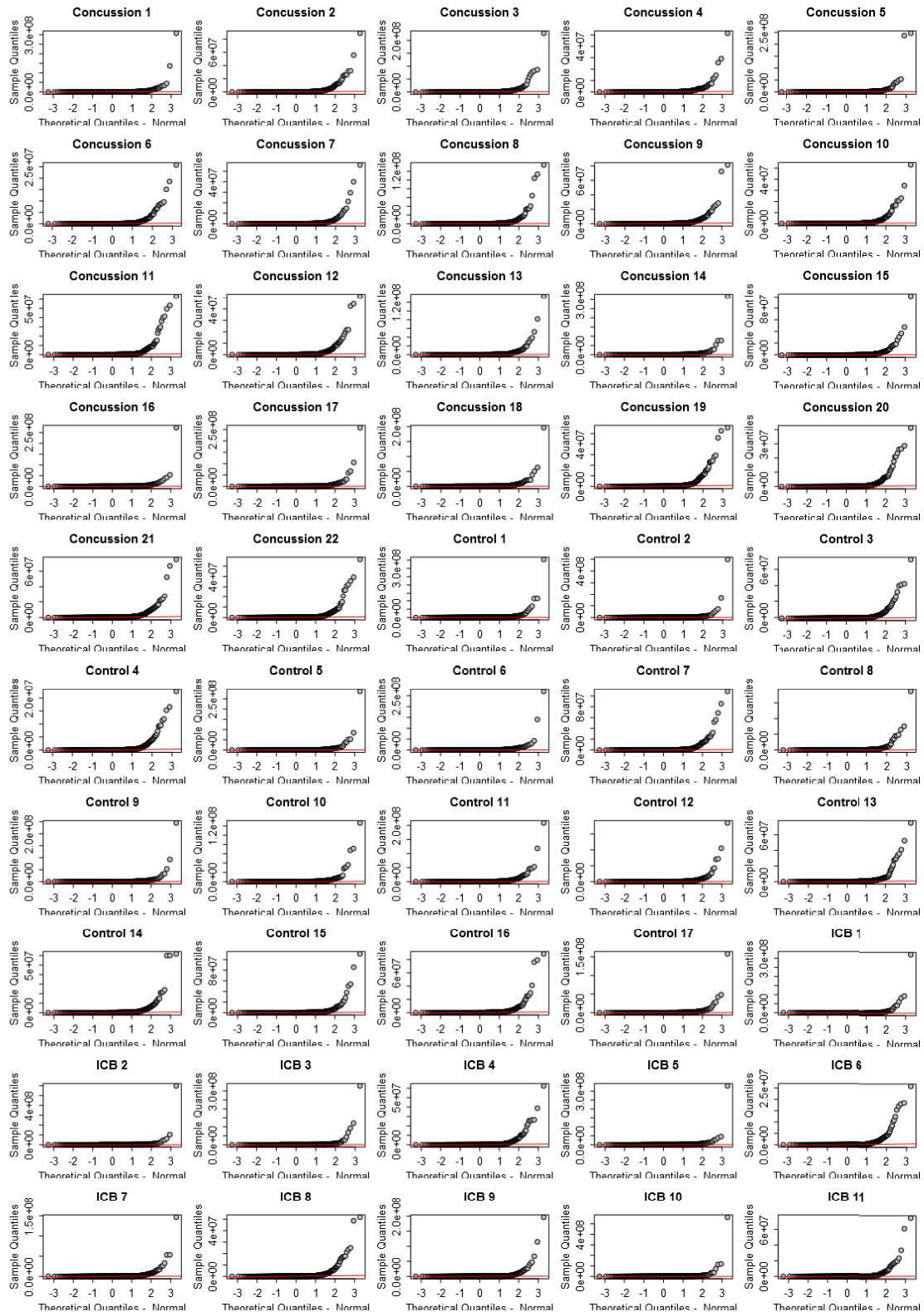


Figure 3.7: Q-Q plot for all 50 DIA analyzed samples. Sample quantiles are plotted against theoretical quantiles with an underlying normal distribution. Once data points form a diagonal, normal distribution for the sample has been proven. Any other graphical distribution provides evidence for non-parametric distribution. Highly abundant proteins as Uromodulin and Serum albumin have been removed from consideration in this analysis. Sample label based on group affiliation – no specific order.

After filtering (such as e.g. removing “reversed” and “cRAP” entries) the list of reported proteins, the protein quantification was characterized via Q-Q-plot estimation. Even though this data was acquired on a different instrumental setup, there was no observable evidence for an existing normal distribution (*cf.* **Figure 3.7**). Hence, protein quantifications have to be handled under consideration of non-parametric distributions. This in particular, allows for the use of the Mann-Whitney u test for investigation of statistical significant protein abundance changes.

3.3.3.3 Biomarker Qualification of TNR16, NGAL and RBP4

In the initial dataset, three diagnostic biomarker candidate proteins were discovered in the context of mild traumatic brain injury. This extended dataset was specifically used to investigate the protein abundance pattern of these three proteins in particular. Nonetheless, the dataset was acquired on a different instrument which can lead to a different protein abundance pattern. However, all of the aforementioned possible biomarker candidates were identified as well as quantified in this dataset. Out of the previously discovered three candidates, only one protein, namely RBP4 (Retinol-binding protein 4) demonstrated a significant abundance change ($p < 0.05$) while comparing the acquired protein abundances between mTBI cases and the corresponding controls (*cf.* **Table 3.4**).

Table 3.4: List of observed p-values in the extended cohort. Different p-values are given: ¥ - mTBI samples compared to only controls (ICB excluded); £ - mTBI samples compare to all control samples (ICB included); β – mTBI samples compared to only ICB samples (ICB explicitly).

Protein Accession	Gene code	p-value ¥	p-value £	p-value β
P08138 (TNR16)	NGFR	0.294677	0.584217	0.731068
P80188 (NGAL)	LCN2	0.681314	1.000000	0.579751
P02753 (RET4)	RBP4	0.027166	0.113405	0.908787

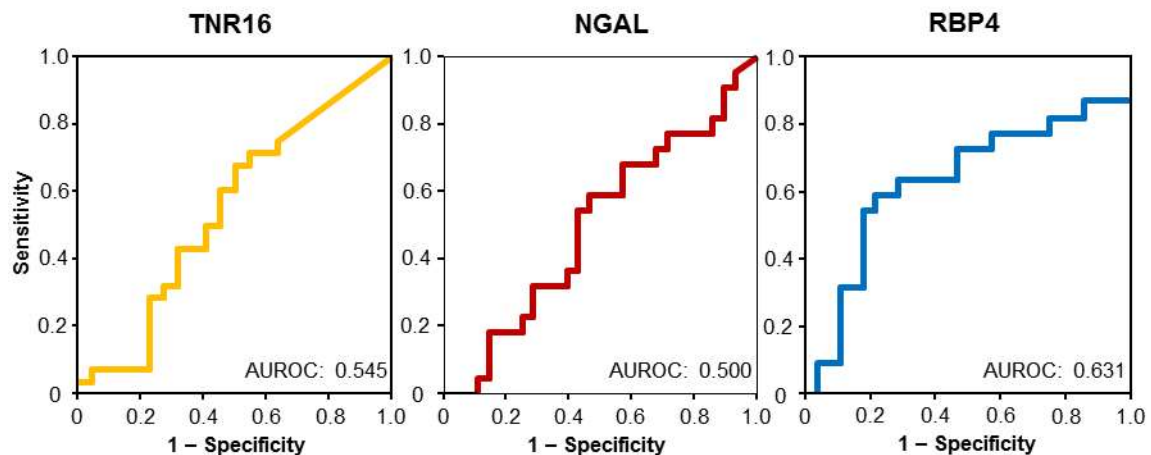


Figure 3.8: ROC curve analysis for identified biomarker candidates. Three different ROC curves for TNR16 (yellow), NGAL (red) and RBP4 (blue) are shown. Ties cause diagonal curve-segments. The corresponding area under the curve (AUROC) demonstrates the diagnostic efficiency for each protein: TNR16 – 0.545; NGAL – 0.5; RBP4 – 0.631. AUROC values in the range of $0.75 < x \geq 0.5$ demonstrate indecisive results.

Additionally, the acquired protein abundances were entered into a ROC analysis to estimate the diagnostic efficiency of these markers in the present dataset. Here, TNR16 scored an AUROC of 0.545, NGAL an AUROC of 0.500 and RBP4 an AUROC of 0.631 (*cf.* **Figure 3.8**). By taking the combined diagnostic effect of all three candidates at the same time (binary logistic regression) an AUROC of 0.503 was achieved. Given these results, it can be concluded that the previously discovered biomarker candidates for mTBI could not be reconfirmed in this extended sample cohort. Hence, these biomarker candidates were rejected, but can still be considered as possible biomarker candidates, which however need further extensive validation.

3.3.4 Secondary mTBI Biomarker Discovery

The DIA acquisition routine can be considered a semi-targeted acquisition methodology. Instead of particularly focusing on one specific precursor ion and its consequent transitions, all precursors out of a specific extraction window are fragmented and the resulting fragment ions are acquired in a MS² spectra. Instead of ignoring valuable information during acquisition, i.e. non-acquired information due to targeted acquisition, the DIA routine allows for the acquisition of additional information of other proteins and still featuring the same information as gained through targeted acquisition. Hence, even though this extended sample cohort was dedicated for the evaluation of the discovered biomarker candidates, the additional information acquired can be exploited for a secondary biomarker discovery investigation.

At first, the protein group abundances were tested for statistical significant changes. Hereto, the abundances of mTBI samples were directly compared to control samples via the Mann-Whitney u test, as non-parametric conditions have to be assumed (*cf.* **Section 3.3.3.2**). This comparison revealed 35 proteins with a p-value of < 0.05. Based on the objective of discovery, multiple testing correction was waived. Besides the comparison of mTBI samples with control samples, mTBI samples were also statistically compared to control samples with ICB samples included. In turn, the second comparison revealed a total of 26 significantly different abundant proteins ($p < 0.05$). Here, only eight protein groups (15.1%) showed significant protein abundance changes in both

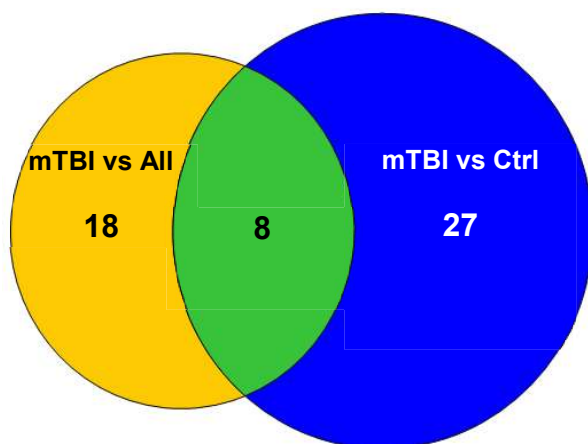


Figure 3.9: Venn diagram for identified candidates. The comparison between mTBI and Controls (Ctrl; blue) demonstrates 35 significant proteins, while the comparison between mTBI and Controls with ICB samples features 26 significant proteins. Eight proteins showed significant changes within both comparisons

comparisons (*cf.* **Figure 3.9**).

In this context and through the specifically chosen sample cohort, an ideal biomarker candidate should be able to distinguish between a healthy state and a concussion, but should also be able to signal once a concussion has already progressed to an intracranial bleed (ICB). Hence, the 26 significantly different abundant proteins which were observed within the comparison of mTBI samples to controls and ICB samples will be further evaluated as potential biomarker candidates in this secondary discovery study.

Consequently, as a first investigation, all protein groups were entered into a PCA analysis. Nonetheless, no clear or even segregation pattern could be deviated from the corresponding PCA plots (cf. **Appendix – Section 8.2.2**). In this PCA analysis, the first three principal components (PC1, PC2 and PC3) inherit the majority of the variation of these identified proteins. However, even in a three-dimensional PCA plot, which further accounts for 63.9 % of the variation within the significant proteins, no segregation pattern can be deviated.

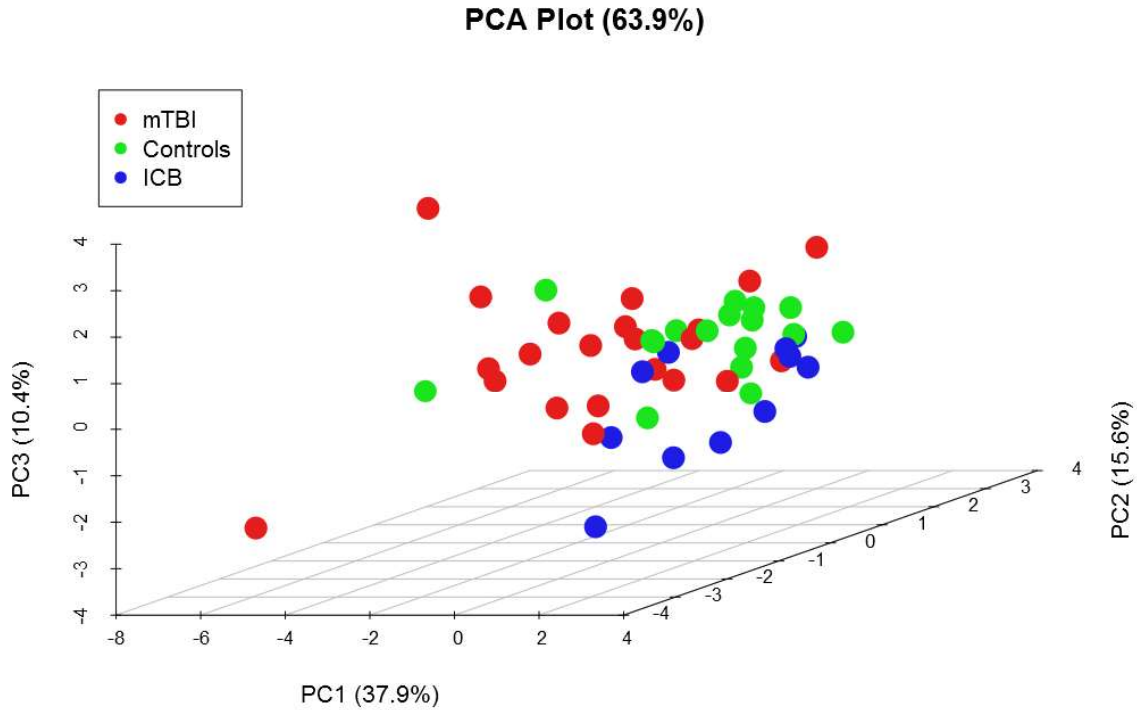


Figure 3.10: 3D PCA plot for the identified 26 significant protein groups. The Top 3 principal components (PC 1 – 37.9 %; PC 2 – 15.6 % and PC 3 – 10.4 %) account for a total dataset variation of 63.9 % within the identified significant protein groups. No sample separation can be deviated

Accordingly, all 26 proteins were entered into a ROC analysis. Here, all proteins featured an AUROC of greater than 0.6. The overall highest scored AUROC was 0.735 for HV313 (P01766, Immunoglobulin heavy variable 3-13). Based on these findings, the top four proteins with the highest AUROC were combined via regression and their combined effect tested for diagnosing ability. Nonetheless, here an AUROC of only 0.849 was scored, which in turn does not represent ideal biomarker conditions (AUROC of 1). Additionally, the parallel use of four proteins to diagnose mTBI, i.e.

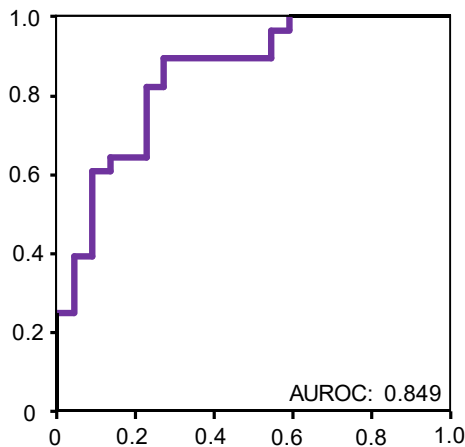


Figure 3.11: ROC of biomarker candidate panel. Analysis of the diagnostic efficiency of the chosen biomarker panel. Area under the ROC curve is 0.849. Perfect condition given with an AUROC of 1.

to segregate cases from controls, but not achieving a fully clear segregation leads to the consequent rejection of the corresponding candidate panel from further consideration.

Based on the aforementioned observations, it can be concluded that statistically significant protein abundance changes can be observed in urine. However, these abundance changes do not necessarily correlate to the existence of a concussion and hence might rather be based on strong interpersonal variances (*cf.* **Appendix 8.2.3**).

3.3.5 Urinary Biomarker Discovery for Intracranial Bleed

In the initial biomarker discovery study, two patients suffering from intracranial bleed (ICB) were entered into the study as positive control. Nonetheless, two samples do not form a representative basis leading to the exclusion of those samples from further consideration in the initial discovery study. However, in the present extended sample cohort, intracranial bleed cases are represented by 11 samples. This indeed forms a representative basis in this study, which in turns allows for the effort of biomarker discovery in the regard of ICB. Hence, instead of mTBI samples being set as “case” samples, now patients suffering from an intracranial bleed were considered as “case” condition.

In general, ICBs can be easily observed and diagnosed through CT or MRI. Nonetheless, the requested time between admission to an ER and the final diagnosis is fairly prolonged, which can hinder first response therapy attempts and hence prolong the patients’ recovery. Thus, it is also of significance to discover urinary based biomarkers for ICB, to quickly, easily and non-invasively determine if a patient is suffering from brain bleedings.

Firstly, the protein abundances of ICB samples were statistically compared (Mann-Whitney u test) to only control samples alone. Here, a total of 124 protein groups were observed with $p < 0.05$. This particular list furthermore features three proteins, that demonstrate a p-value of < 0.001

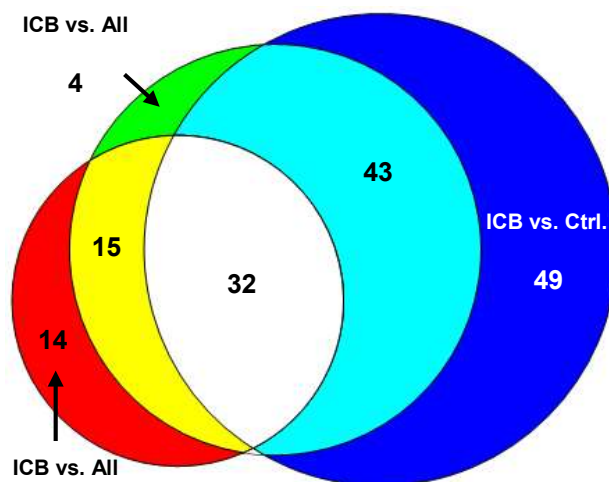


Figure 3.12: 3-way Venn-diagram for all comparisons. Diagram is based on protein groups observed with statistical significant protein groups. The absolute intersection features 32 protein groups that shows significance ($p < 0.05$) in all three comparisons: ICB samples vs. all control samples (green), ICB samples vs. only negative controls (blue) and ICB samples vs. only mTBI samples (red)

(Q8IUK5, Plexin domain-containing protein 1; Q8IYS5, Osteoclast-associated immunoglobulin-like receptor; P00167, Cytochrome b5) even featured a p-value of < 0.001 . Consequently, the mTBI samples were incorporated into the control group and hence the extended “control group” compared to the ICB group. This comparison revealed a list of 94 proteins demonstrating a statistically significant protein abundance change ($p < 0.05$). Here as well, three samples showed significance of $p < 0.001$. In this case, two proteins are identical to the former comparison: Q8IYS5, Q8IUK5. Nonetheless, also for this comparison a multiple testing correction was waived.

Even so it was hypothesized that an ideal biomarker candidate should be able to distinguish between either control alone or also between control and mTBI. Hence, for further analysis only

proteins demonstrating a significant protein abundance change in both comparisons were chosen. The resulting intersection between both comparisons consists of 75 proteins. Additionally, the comparison of ICB sample directly to mTBI samples, revealed 61 protein groups, which showed significance of $p < 0.05$. The 3-way Venn-diagram intersection features a total of 32 protein groups, which consequently feature a significant ($p < 0.05$) protein abundance change in all three comparisons.

In this discovery study, proteins do not necessarily feature a significant protein abundance change while comparing ICB samples vs mTBI samples. Hence, as a consequent analysis, all 75 proteins within the intersection of the comparisons of ICB samples vs. all control samples (negative controls and mTBI samples) and ICB samples vs. only negative control samples were entered into a PCA analysis. In this analysis, the first principal component (PC1) already accounts for 55.2 % of the variance of the entered proteins. Further components (PC2 and PC3) account for only either 7.51 % or 6.22 %, respectively (*cf. Figure 3.13*). Based on this fact, only component combinations with PC1 are considered valuable (*cf. Appendix 8.2.4*). Nonetheless, a clear group-wise segregation of controls with mTBI samples and ICB samples was not achieved. However, a certain segregation pattern can be deviated. Here, ICB samples tend to collectively segregate towards the right end of the corresponding PCA plots. Furthermore, based on the consequent PCA Bi-plots, no feature was detected that actually causes the ICB-specific segregation effect. It appears that ICB samples simply behave in a reciprocal proportional manner to the corresponding control samples.

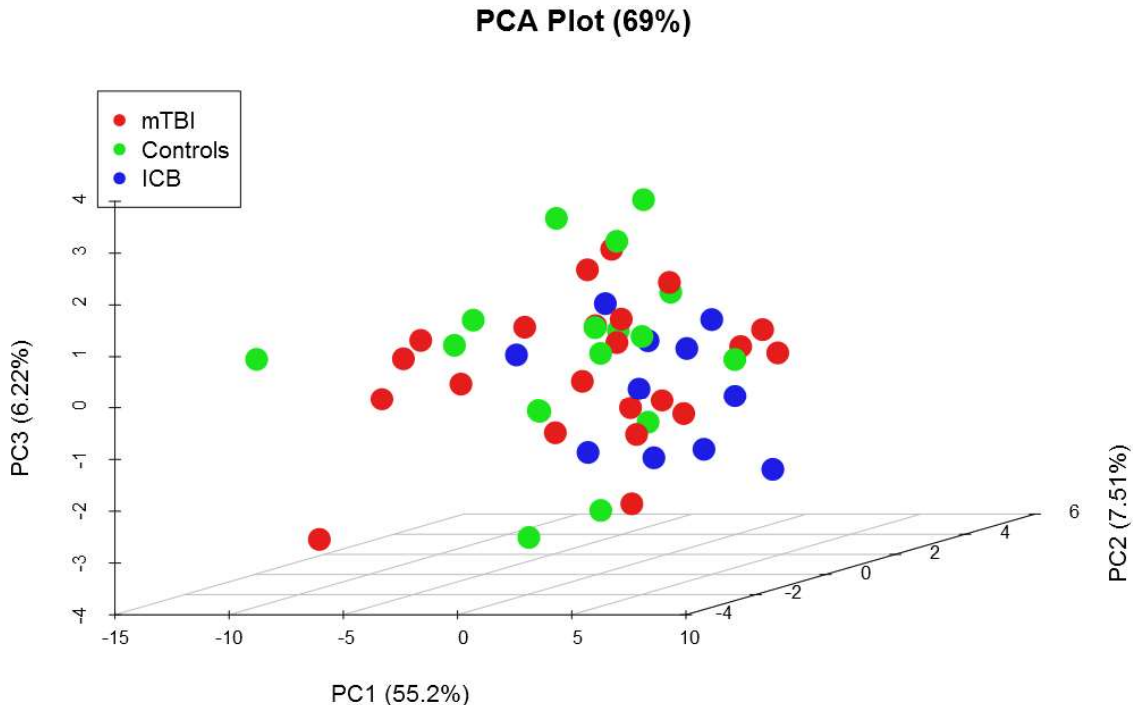


Figure 3.13: 3D PCA plot for all 75 significant protein groups. PCA plot considers the first three principal components: PC 1, PC 2 and PC 3. The plot accounts for a total of 69 % of the dataset variation. Here, the first component (PC 1) already accounts for 55.2 % of the variation. Consequently, PC 2 accounts for 7.51 % and PC 3 for 6.22 % of the variance. The plot shows no group-specific segregation.

Since no feature, or i.e. protein, causes an ICB specific segregation tendency, all previously mentioned 75 proteins were further investigated in the regard of ROC curve analysis and its corresponding area under the curve (AUROC). Each protein alone scored an AUROC of ≥ 0.69 . This indeed signals that there is a diagnosing tendency for ICB, however, the diagnosing effect of ultimately a single protein is not present. The overall highest AUROC (0.853) was scored by O75339 (Cartilage intermediate layer protein 1) which also demonstrates one of the lower p-values ($p < 0.01$) in both comparisons. In turn, the aforementioned protein demonstrated the lowest p-value in the comparison of ICB samples vs. all control samples (negative controls and mTBI samples). Consequently, all proteins with an AUROC of ≥ 0.8 were combined into a biomarker candidate panel and also tested for diagnosing efficiency.

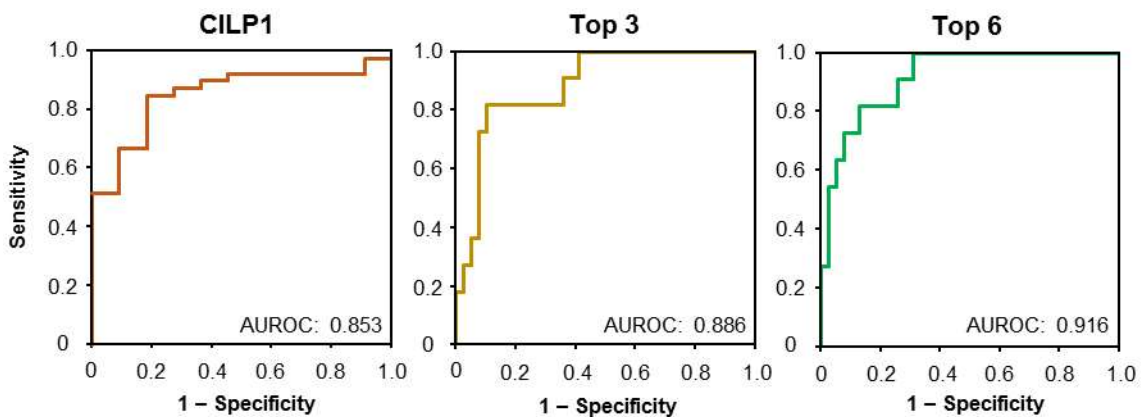


Figure 3.14: ROC curves for ICB diagnosing efficiency analysis. The analysis of single proteins revealed CILP1 (P30622; orange) with the highest AUROC of 0.853. Furthermore, the top three AUROC scoring proteins were combined into a biomarker panel (yellow) and tested for diagnostic efficiency. An AUROC of 0.886 was scored. Lastly, all proteins with an AUROC of ≥ 0.8 were linearly combined and tested (green). This resulted in an AUROC of 0.916.

Here, a total of two different combinations was tested: the top 3 proteins only (O75339, Q81Y55 and Q81VK5) and Top 6 (O75339, Q81Y55, Q81VK5, Q9UBD6, O60494, P11047 and P16035). With the combination of the top six proteins, all those proteins with an AUROC of ≥ 0.8 were considered for a possible assignment to the biomarker panel (cf. **Figure 3.14**). Nonetheless, no clear diagnostic event was achieved even by combining 10 proteins into the panel. Here, the AUROC was optimized until it reached 0.916 but could not be improved any further. The corresponding usage of these six proteins, allows for an 88 % correct prediction of the patients' state (cf. **Appendix - Figure 8.2.9**).

Given the rather low percentage of correct prediction and the insignificant abundance changes after a multiple testing correction (e.g. Benjamini-Hochberg or Bonferroni) it can be deviated that the diagnosing tendencies for ICB in this dataset may rely on a random event and thus won't be able to be reconfirmed in a follow-up study. Hence, it can be concluded that ICB cannot be predicted simply by protein abundance changes in the urine of injured patients.

3.4 Conclusion and Perspective

Concussion (mTBI) or Intracranial bleed (ICB) are imposing mild to severe brain injuries. In this regard, a quick diagnosis is inalienable to initiate an early therapy, particularly in the prevention prolonged recovery or secondary brain damage. Nonetheless, to this day there is no objective measurable molecule or biomarker that allows the diagnosis of concussion. Furthermore, intracranial bleeds can only be diagnosed after CT or MRI and a brain lesion hereby observed. Consequently, physicians are in desperate need of quick, objectively and non-invasively measurable events to diagnose these diseases.

Urine itself is produced by the kidneys and so, is based on the filtration of the blood stream in the glomeruli. By its definition, urine is considered the waste fluid of the body and can be obtained easily, in vigorous amounts and without any invasive methodology. Hence, urine portrays the perfect body fluid for the use of diagnosis. In this study, the scope was on the discovery of neuronal damage related processes and hence biomarkers in urine. However, a potential biomarker or biomarkers would first need to pass the blood-brain barrier to reach the glomeruli to then be filtered and secreted into the urine. It is also discussed, that urine and/or saliva as well as tears might not be the body fluid of choice for the discovery of neuronal associated diseases, as they are considered too far away from the scene of where disease-specific processes are taking place¹³⁸. Additionally, kidney-related processes and their corresponding by-products can cover the actual low-abundant neuronal biomarkers. Hence, it can be concluded that there still might be biomarkers in urine with regard to neuronal damage, but that these markers are locked in the extensive dynamic range of proteins in urine and cannot yet be revealed by current technological equipment.

3.5 Acknowledgments and Contribution

I want to thank Dr. Hanno Steen for insightful discussions and scientific support while working on this project. Additionally, I thank Dr. Jan Muntel for his input for the optimization of the LC-MS/MS analysis of urinary samples in this study. Also, I want to thank Dr. Rebekah Mannix and Dr. Mark Berry for the establishment of the urinary sample cohort and all various research coordinators for the collection of urine samples in the Emergency Medicine Department at Boston Children's Hospital (Boston, MA, USA).

My personal contributions to this project account for: sample management, Standard Operating Procedure (SOP) establishment, carrying out the described experiments, analyzing the samples by mass spectrometry, data analysis and figure generation.

4. Exosomal Biomarker Discovery for mTBI in Urine

4.1 Introduction

Urine is considered a waste fluid of the human body ¹⁶³. It is a by-product of the kidneys and is specifically used to secrete useless metabolites and other undesirable substances. A healthy human individual secretes approximately 1.2 l per day, which can vary between 0.6 to 2.4 l based on hydration, activity level, environmental factors and weight ¹⁶⁴. Even though urine is considered waste fluid, it is increasingly making its way into the scope of research as valuable information, especially regarding kidney and systemic malfunctions can be non-invasively obtained ¹⁶⁵⁻¹⁶⁷.

The kidneys filter the blood and rejects unwanted molecules and substances which have no additional benefit or may even harm the human body into urine to be secreted. In this context, renal epithelial cells in particular, secrete nanovesicles (40 -100 nm) with endocytotic origin, into urine, especially when a multivesicular body (MVB) fuses with the apical plasma membrane ^{168, 169}. These secreted nanovesicles are hence also called “exosomes”.

Urinary exosomes account for approx. 3% of the whole urinary proteome ^{170, 171} and can provide a particular insight into the renal physiology ¹⁷². Nonetheless, exosomes, per se, can also inherit a signaling function and hence can change a recipient’s proteome via its signaling feature. Since these nanovesicles are non-invasively sustainable and contain not only proteins, but also RNA and metabolites ¹⁷², exosomes are of particular interest ¹⁷³.

These facts make exosomes a very interesting research target and might help to explain certain disease-related processes or even in the discovery of new disease-related biomarkers. However, due to the extended preparation and purifying process, the actual clinical applicability still has to be proven, as well as the clinical information content remaining questionable ¹⁶⁹.

In this study, exclusively the urinary exosome proteome is further investigated with the scope of discovering processes or proteins, which are then related to sustaining a concussion or intracranial bleed.

4.2 Experimental Procedures

4.2.1 Urine Sample Collection

Urine samples were collected from consenting patients presenting to the ED at Boston Children's Hospital within 24 hours after concussive injury. After collection, the sample was directly transferred to the research facility by the associated research coordinators. Upon receipt the samples were aliquoted (depending on sample volume: 10x 2 ml; 1x 15 ml; 1x 25 ml and residual urine into 50 ml aliquots) and consequently frozen at -80°C . This study was reviewed and approved by the Institutional Review Board (IRB) at Boston Children's Hospital. Only samples with sufficient sample volume (i.e. at least one 25 ml aliquot) were entered into this exosome study.

4.2.2 Urinary Exosome Extraction

Firstly, urine (5 or 10 ml) was thawed in a standardized water bath at 25°C and kept in bath no longer than 15 minutes. Samples which were not completely thawed were kept on RT until being completely liquid. For further processing, samples were kept on ice to avoid any spontaneous degradation. Cellular debris and other solid impurities were removed by pelleting at 4,000 rpm for 10 min. Supernatant was carefully removed, not to disrupt the pellet's integrity and transferred to an ultra-centrifugation (UC) tube (max. vol. 10 ml). Depending on the extraction volume used (5 or 10 ml), the absolute volume (10 ml) needed was achieved by adding water (HPLC grade). Imbalance was avoided by balancing the weight for each centrifugation tube. UC was performed via a swing-bucket rotor (SW 41 Ti, Beckmann Coulter) in a vacuumed chamber ultra-centrifuge (Optima L-90K, Beckmann Coulter) at 40,500 rpm (approx. $200,000\times g$) for 1 h. Supernatant was carefully removed so not to disrupt the pellets integrity and consequently stored at -20°C . Exosomes were washed by re-suspending the pellet in 5 ml 1x Phosphate-buffered saline (PBS, pH 7.4) and consequently re-pelleted by UC at approx. $200,000\times g$ ($70,000$ rpm) for 1 h in a TLA-100.3 rotor (Beckmann Coulter) and Optima™ MAX-E Ultracentrifuge (Beckmann Coulter). Supernatant was removed and discarded. The residual exosome pellet (slightly yellow but clear) was submerged in urea supernatant (~ 8.3 M urea solution) and incubated for an hour at RT on a Petri dish shaker (OrbiShaker, Benchmark) with max. speed settings.

4.2.3 Filter-assisted sample preparation (FASP)

The urea-exosome suspension (*cf.* **Section 4.2.2**) was supplemented with 30 μl DTT (100 mM DTT in 1 M Tris/HCl pH 8.5), transferred to low protein adsorbing microcentrifugation tubes (LoBind, Eppendorf) and incubated for 20 min at 27°C and 1100 rpm in a thermo mixer. Reduced cysteine side chains were alkylated with 50 mM IAA (final concentration) and incubation for 20 min in the dark at 27°C and 750 rpm. Fully denatured proteins were captured on a 10 kDa MWCO spin filter (MRCPR010, Millipore) and washed twice with 50 mM ABC. Protein digestion was performed with sequencing grade trypsin (V5111, Promega) by adding 1 μg total trypsin in 100 μl 50 mM ABC to each sample and overnight incubation at 37°C in a humidified incubator. Resulting peptide species were eluted with 300 μl 0.5 M NaCl. For further analysis, peptide eluate was divided into two equal aliquots.

One aliquot was desalted with reversed phase-based TARGA C-18 spin tips (SEMSS18R, Nest Group) for DIA based LC-MS/MS analysis. Lyophilized samples were stored at -20°C for further analysis. The other aliquot was either directly fractionated (5 ml cohort) or pooled (10 ml cohort), according to its sample group affiliation, desalted and consequently fractionated via reversed phase-based graphite carbon spin tips (TT2CAR, GlySci) for high ion library depth and hence DDA based LC-MS/MS analysis.

4.2.4 Graphite Carbon Peptide Fractionation

FASP-based peptide eluate pools (*cf.* **Section 4.2.3**) were acidified with 50 µl 20% trifluoroacetic acid (TFA) and quickly vortexed to achieve a homogenous acidification. Graphite carbon tips were primed by applying 200 µl of 95% ACN with 0.1% TFA and pressurizing the tip via in-house pressurized air supply. This step was repeated once. Consequently, the tip was preconditioned twice by applying 200 µl 0.1% TFA in HPLC grade H₂O. This step also was repeated once. Each flow-through was discarded. Acidified sample (~200 µl) was applied and pushed through the graphite carbon bed. For enhanced binding efficiency, the sample binding step was repeated once with corresponding flow-through. The sample was consequently desalted by applying thrice 200 µl of 0.1% TFA in HPLC grade H₂O. The flow-through was discarded. For fractionation, peptides were stepwise eluted by increasing the acetonitrile content in the elution buffer. Here, a total of 8 fractions was achieved: 20, 25, 30, 35, 40, 60 and 95% ACN. Enhanced fractionation efficiency was achieved by eluting twice with ACN containing buffer and once with 0.1% TFA. All correlating flow-throughs were combined in one microcentrifugation tube. Based on the reversed phase-based fractionation mode as well to shorten needed DDA acquisition time, four fractions were combined: 20 % fraction with 95 % and 25 % with the 60 % fraction. Samples were lyophilized and stored at -20°C until further analysis.

4.2.4 LC-MS/MS Analysis

4.2.4.1 Data-dependent Acquisition for a fractionated ion library (fLib)

The six peptide fractions were reconstituted in loading buffer (5% ACN (v/v); 5% FA (v/v)) and iRT calibration peptides spiked in a 1:30 ratio (1 µl peptide stock in 29 µl sample). Trypsin digests (~1.2 µg; NanoDrop normalized) were directly injected onto an in-house packed (ReproSil-Pur C18 3 µm, Dr. Maisch GmbH; 15 cm) fused silica column (75 µm x 300 mm, New Objective) and peptides were consequently separated by a linear gradient from 93 % buffer A (0.2 % FA in HPLC water; buffer B: 0.2 % FA in ACN) to 70 % within 60 min. Analytes were ionized and sprayed into the mass spectrometer (Q Exactive, Thermo Fisher) by applying 2.2 kV to the spray emitter. The mass spectrometer was operated in data-dependent TOP10 mode with the following settings: mass range 250 – 1800 Th; resolution for MS1 scan 70 000 @ 200 Th; lock mass: 445.120025 Th; resolution for MS2 scan 17 500 @ 200 Th; max. injection time (IT) 120 ms; isolation width 1.6 Th; Normalized Collision Energy (NCE) 27; underfill ratio 1.0 %; charge state exclusion: unassigned, 1, >6; dynamic exclusion 30 s.

4.2.4.2 Data-independent Acquisition for quantitative analysis

Lyophilized sample peptide mixtures were reconstituted in loading buffer (5% ACN (v/v); 5% FA (v/v)) and peptides for hyper-reaction monitoring (HRM) calibration peptides were spiked-in with a 1:30 ratio (1 µl peptide stock in 29 µl sample). Tryptic digests (~1.2 µg; NanoDrop normalized) were directly injected onto an in-house packed (ReproSil-Pur C18 3 µm, Dr. Maisch GmbH; 15 cm) fused silica column (75 µm x 300 mm, New Objective). Peptide mixtures were separated by a linear gradient from 93 % buffer A (0.2 % FA in HPLC water; buffer B: 0.2 % FA in ACN) to 70 % buffer A within 60 min. Analytes were ionized and sprayed into the mass spectrometer (Q Exactive, Thermo Fisher) by applying 2.2 kV to the spray emitter. DIA cycle contained one full scan (MS¹) and 17 consecutive parallel reaction monitoring (PRM) scans covering a mass range of 400 – 1040 Th. The full scan resolution was set to 35,000 @ 200 Th; AGC target –3e6, maximal IT 120 ms; mass range 400 – 1,200 Th; followed by DIA scans with resolution 35,000 @ 200 Th; variable window size was applied; NCE 27; AGC target value 1e6, max. IT 120 ms. Extraction windows are given in **Tabel 4.1**.

Table 4.1: Extraction windows applied for DIA analysis.

Window No.	Window size
1	26 Da @ 413 m/z
2 to 8	27 Da @ 437.5, 462.5, 487.5, 512.5, 537.5, 562.5 and 587.5 m/z
9 to 12	32 Da @ 615, 645, 675 and 705 m/z
13 to 16	52 Da @ 745, 795, 845 and 895 m/z
17	81 Da @ 959 m/z

4.2.5 Data Analysis

4.2.5.1 DIA Analysis

The fLib was generated with MaxQuant ¹¹⁷ (v.1.5.3.30) via the implemented Andromeda algorithm. The acquired pooled and fractionated DDA files (*.raw) were loaded into MaxQuant and consequently searched against a reference protein sequence database, here, the human UniProtKB database (Homo sapiens, ~20,300 sequences, version 06-2014) was used with “Commonly occurring laboratory contamination protein” sequences (cRAP, version 2012.01.01) added. Variable Modifications were set to: Acetyl(Protein N-Term), Oxidation(M), Deamidation(NQ) and Phospho(STY). For fixed modifications only Carbamidomethyl(C) was used. The digestion mode was set to “Semispecific” and the digesting enzyme to “Trypsin”. Feature matching across runs was activated with a 0.7 min matching window and a 20 min alignment window, as acquired files are based on fractionated samples. Otherwise default settings (by MaxQuant) were used for the analysis: trypsin with up to three missed cleavages; mass tolerance for the first search: 20 ppm; main search: 4.5 ppm.

Corresponding DIA files (*.raw) were fully processed and analyzed in Spectronaut ¹⁴⁷ (v.9.0.1 (Collins); Biognosys). At first, the search results from the aforementioned MaxQuant search of

the corresponding DDA acquisition were loaded into Spectronaut to import the ions for the fLib, which in turn was saved within the program's data-structure. The ion library import was performed via the default settings, proposed by Spectronaut. Nonetheless, peptide and protein grouping was kept in the results by the Andromeda search engine (MaxQuant). Hence, the "Protein Inference" option was deactivated for ion library generation. After generating the ion library, corresponding DIA files were loaded into Spectronaut and the aforementioned ion library was assigned for analysis. The following custom settings were applied in Spectronaut: Calibration Mode: Forced; Protein Quantity: Sum Precursor Quantity; Data Filtering: Either QValue or QValue Sparse. All results were filtered by a Q value of 0.01 (equals a FDR of 1% on peptide level). The output report was adjusted to contain all quantification related values. All other settings were kept on default.

4.2.5.2 Statistical Analysis

The Spectronaut™ output was statistically analyzed with either SPSS v23 (IBM) or EXCEL with "Real Statistics Resource Pack" plugin. Further statistical analysis was performed with the proteomics data analysis tool: Inferno (v.1.1.6044) ¹⁷⁴ from <http://omics.pnl.gov>.

4.3 Results and Discussion

4.3.1 Sample Cohorts

The extraction of urinary exosomes out of high-volume samples has already been described^{170, 175-177}. Samples used in this study are however limited in quantity. Furthermore, the exosomes correlating to one particular sample and patient were dedicated for investigation and not as a previously described pooled sample. Here, two sample cohorts were used to investigate urinary exosomes. The data of the initial dataset is based on exosomes, extracted out of 5 ml sample urine. The follow-up exosome study is based on exosomes extracted out of 10 ml of urine.

Within the first exosome extraction, only seven samples out of the concussion group were used for extraction. Additionally, nine samples out of the control group and five samples from intracranial bleed patients could be entered into the study. Retrospectively, only one concussion sample, three control samples and one intracranial bleed case had to be excluded due to insufficient sample urine. In total, exosomes from 21 samples were extracted, lysed and analyzed with mass spectrometry.

Subsequently, the exosome proteome was reviewed by an extended sample cohort and using twice as much starting material as in the former study. Here, a total of 34 samples were entered into the study; 12 concussion samples, 15 control samples and 7 intracranial bleed samples were included in this study.

4.3.2 Ion Library and Protein Identifications

The ion library, or fractionated ion library (fLib) to be specific, is based on sample specific fractionated peptide mixtures for the initial cohort (5 ml sample volume). Samples for fractionation were selected based on their group affiliation (4 x concussion; 3 x control and 4 x ICB), however, the group-specific samples were chosen randomly. After import of the MaxQuant search results into Spectronaut, the ion library featured 2079 unique protein groups, which in turn account for

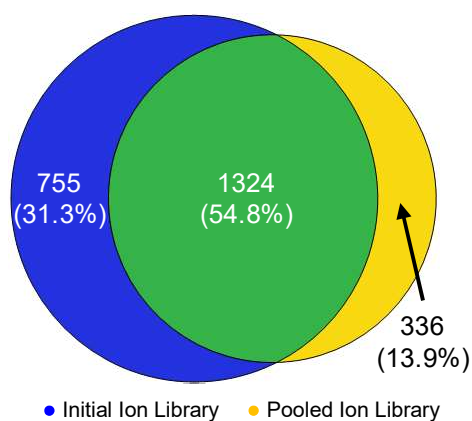


Figure 4.1: Ion Library Comparison. The initial ion library (blue) is based on fractionated single samples, while the pooled ion library (yellow) is based on samples which are pooled and then fractionated. More than the half (54.8%) of the identified proteins is present in both libraries.

2370 unique proteins. Here, 568 protein groups were only based on a single peptide precursor. These “single hit” protein groups represent critical features as their quantitative value might be compromised due to incorrect peak assignments. Nonetheless, single hits were not excluded for this analysis.

In contrast, the ion library for the extended sample cohort is based on fractionated samples which in turn originate from pooling. Here, samples were pooled group-wise into two samples representing each group. Fractionation was performed as for the previous cohort. The import of the search results from MaxQuant resulted in an fLib featuring 1660 unique protein groups, accounting for 1965

unique proteins. Compared to the previous generated ion library, only 186 proteins are represented here by only one peptide precursor, as well as single hits not being excluded for the quantitative analysis.

In comparison, 54.8% of all protein identifications are shared in-between both ion libraries. Furthermore, the initial ion library (fractionated single samples) features additional 755 proteins (31.3% of the protein identifications), while the pooled and fractionated ion library features 336 (13.9%) additional unique protein identifications (*cf.* **Figure 4.1**). Gene Ontology (GO) annotation for cellular compartments of all protein accessions in the ion libraries revealed that in both cases more than 50% (52.84 % of the initial ion library and 63.66% of protein accessions in the pooled ion library) are affiliated with exosomal vesicles (exosomes). As protein accessions can be affiliated to more than one category, the overall percentage exceeds 100% (*cf.* **Figure 4.2**). Based on this annotation, it can be concluded, that the exosome extraction via ultra-centrifugation (UC) was successful. The annotation was performed with the open-source available tool “FunRich”¹⁷⁸ which uses the online available ExoCarta compendium for exosomal vesicle annotation. Here, researchers can upload their generated data sets based on exosomal studies to allow for a full exosomal mapping. However, in this regard contaminating and highly abundant proteins (e.g. uromodulin and albumin) are also listed as “exosomal proteins”. Nonetheless, samples were additionally investigated for the presence of specific exosome surface markers.

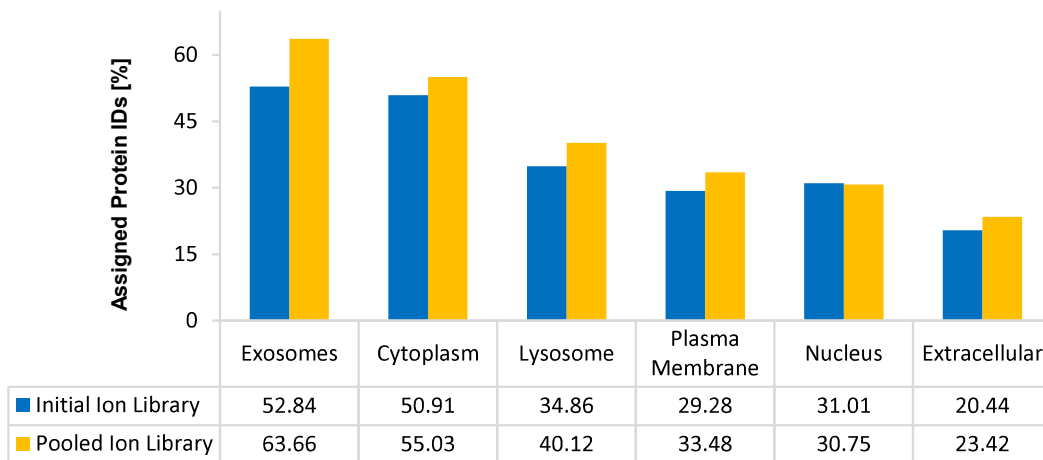


Figure 4.2: GO Annotation for both ion libraries. Gene ontology (GO) annotation for cellular compartments (CC) of the protein accessions identified in each ion library. The pooled ion library features ~10% more protein identifications which are associated with exosomes. In this context, contaminating proteins, which originate from the regular urinary proteome, can compromise the percentage. GO annotation was performed via FunRich¹⁷⁸ and the underlying ExoCarta Exosomal Protein Database¹⁷⁹⁻¹⁸¹.

4.3.3 mTBI Biomarker Discovery in the Urinary Exosome Proteome

Across all analyzed 21 sample DIA files, a total of 1707 protein groups were quantified out of 2079 identified protein groups in the ion library. This in turn, means that 82.1% of the ion library was identified in the DIA files. On average 804 ± 228 proteins were quantified per sample. A maximum of 1257 proteins were quantified in one sample, while a minimum of only 332 proteins were quantified. Non-quantified proteins consequently result in “missing values”, which can impact on statistical analysis. In this dataset, each sample featured on average is 47.2 ± 13.3 % of

missing values. In one case the number of missing values even rose above 80%. Nonetheless, samples with an even higher number of missing values were not excluded. In **Figure 4.3** the missing value distribution is demonstrated.

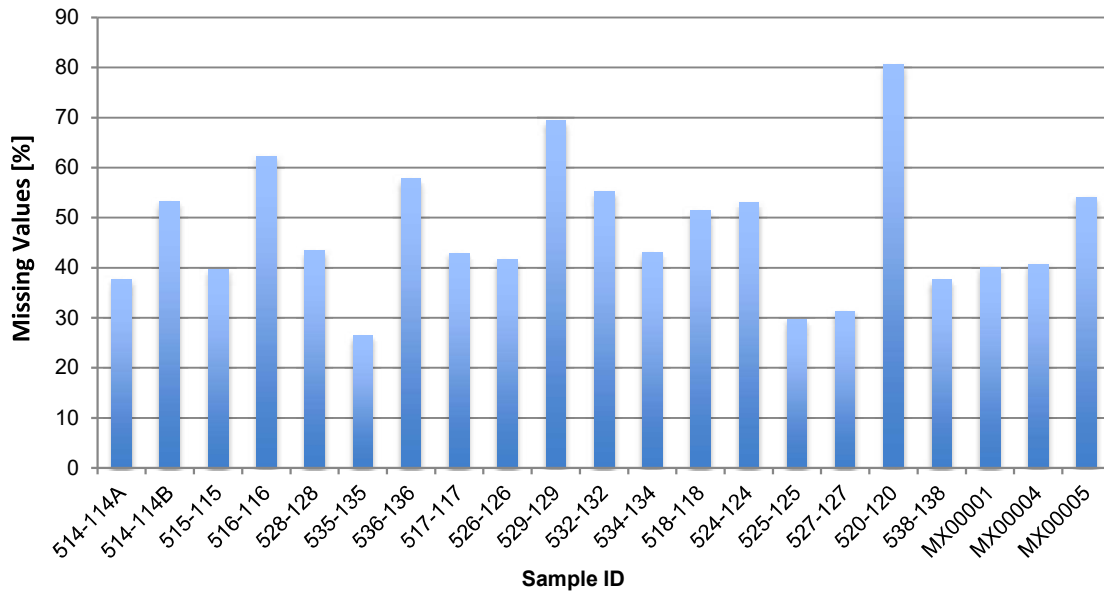


Figure 4.3: Extend of missing values per sample. Missing values can inherit a negative impact on the statistical analysis. In this dataset, the amount of missing values varies between 26.4 and 80.5%. Samples were not excluded from analysis solely based on missing values.

After sum normalization the first PCA was performed. Here, the PCA plot does not demonstrate any clear group-based separation. Nonetheless, it was observed that one sample (520-120; ICB) always demonstrated outlier characteristics. In turn, this sample also demonstrated the most missing values in this dataset. Based on the assumption that the amount of missing values correlates with an unsuccessful exosome extraction, this sample was excluded from further considerations. However, the exclusion did not change the outcome of a PCA and still, no obvious group-specific separation was observed (*cf. Appendix 8.3.1 ff.*).

For a subsequent statistical analysis for significant protein abundance changes, the sample specific quantitative values were first tested for normal distribution via Q-Q-plots. Nonetheless, a normal distribution was not confirmed as no Q-Q-plot demonstrated a diagonal line (*cf. Appendix 8.3.1*). Normal distribution was not confirmed for any sample, which requires non-parametric data handling for further analysis. Consequently, protein abundance values were compared according to their grouping for statistical significance via Mann-Whitney u test. Based on the study setup, three distinct groups are available: Concussion group (mTBI), Control group (CT) and intracranial bleed group (ICB). The comparison was performed by either comparing mTBI versus Control and ICB group or mTBI versus only Control group. The comparison between mTBI and CT with ICB, a total of 21 proteins demonstrated statistical significant ($p < 0.05$) abundance variations. The comparison of mTBI vs. only Controls, featured a total of 16 significant proteins. Here, the scope of research laid within the discovery of diagnostic biomarkers. Hence, the abundance changes of proteins should be significant for both comparison: mTBI vs. both CT and ICB as well as mTBI vs. only CT.

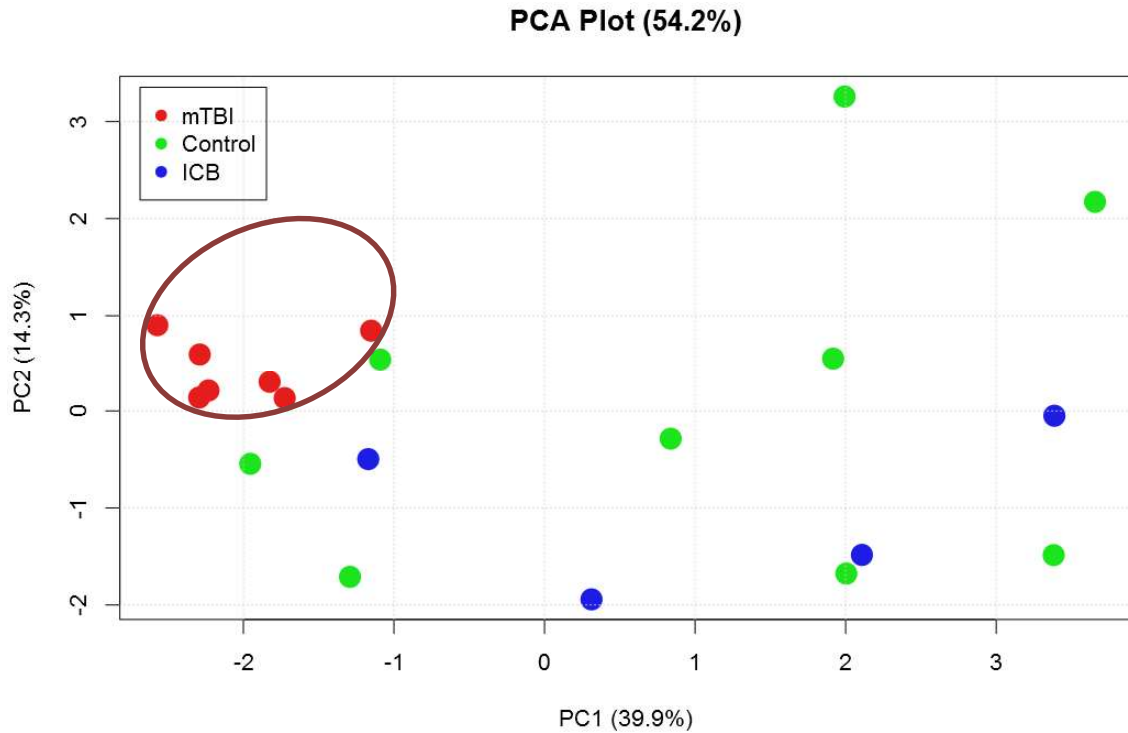


Figure 4.4: PCA plot of proteins with significantly different protein abundance. Group-based clustering of mTBI cases (red circle). Clear separation and classification of mTBI cases from control and ICB samples. Proteins entered into PCA demonstrate classification ability and might function as biomarker candidates.

After filtering proteins based on their statistical significance for protein abundance ($p < 0.05$) changes, another PCA was performed, but due to too many missing, the PCA failed. Circumventing this issue, missing values were then imputed by assigning half of the actual minimal observed abundance (half minimum; HM)¹⁸² level. The corresponding PCA plot (*cf.* **Figure 4.4**) features clear separation between mTBI and Control as well as to ICB samples. This outcome confirms that these exosomal proteins might feature diagnosing power for mTBI with their abundance change in urinary exosomes.

Table 4.2: List of filtered proteins with significant abundance changes.

ID	Accession	Protein Name	p-value
1	Q16560	U11/U12 small nuclear ribonucleoprotein	0.002281
2	A6H8Y1	Transcription factor TFIIIB component B" homolog	0.003823
3	Q15286	Ras-related protein Rab-35	0.005545
4	O75363	Breast carcinoma-amplified sequence 1	0.012554
5	Q9NR99	Matrix-remodeling-associated protein 5	0.012554
6	Q9Y520	Protein PRRC2C	0.012554
7	Q92896	Golgi apparatus protein 1	0.012554
8	Q969S3	Zinc finger protein 622	0.015652

9	P22033	Methylmalonyl-CoA mutase, mitochondrial	0.023919
10	Q9H190	Syntenin-2	0.023919
11	Q3SXY8	ADP-ribosylation factor-like protein 13B	0.043311
12	Q8WZ75	Roundabout homolog 4	0.047583

As final evaluation, the AUROC was determined for each significantly different observed protein alone. In this regard, the ROC curve was calculated by either focusing on detection of mTBI (rule-in) but also on detecting controls (rule-out) amongst controls. Perfect biomarker conditions are given once a single protein scores an AUROC of 1. Additionally, an AUROC of 0.5 reflects a 50% chance of achieving a true positive diagnosis. Hence, in this analysis, AUROCs of > 0.7 were considered possible biomarker candidates. On the one hand, while focusing on identifying rule-in markers for mTBI via AUROC, three proteins demonstrated an AUROC of > 0.7, namely, Q9H190 (0.912), P22033 (0.813) and O75363 (0.769). On the other hand, three proteins also featured an AUROC of > 0.7 while investigating options for rule-out markers: A6H8Y1 (0.846), Q9Y520 (0.780) and Q8WZ75 (0.703).

Nonetheless, a biomarker is not necessarily represented by only a single feature, i.e. protein. The diagnosing power can also be shared between multiple features which consequently form a biomarker panel¹⁸³. In this context, which combination of features can simulate perfect biomarker conditions was evaluated. To minimize the combinations to be tested, only those proteins which demonstrated the same effect were combined (rule-in or rule-out). Here, one combination was identified, which featured an AUROC of 1. Combining the protein abundance information in urinary exosomes of Q9H190 and P22033 led to an 100% correct group prediction, simply by considering the protein abundance of those two proteins. Even though a perfect combination of proteins in the urinary exosome proteome has been identified, which is in turn able to objectively diagnose mTBI based on their protein abundance level, it is wise not to extensively filter possible biomarker candidates before entering the qualification step.

Table. 4.3: List of protein combinations and resulting AUROC.

No.	Protein Accession	AUROC
1	Q9H190 and P22033	1.000
2	Q9H190 and O75363	0.912
3	P22033 and O75363	0.835
4	Q9H190, P22033 and O75363	0.835
5	A6H8Y1 and Q9Y520	0.901
6	A6H8Y1 and Q8WZ75	0.824
7	Q9Y520 and Q8WZ75	0.802
8	A6H8Y1, Q9Y520 and Q8WZ75	0.901

4.3.4 Extended Sample Cohort for Biomarker Qualification

The extended qualification cohort consists of 34 samples which were acquired via DIA. Out of the 1660 protein groups which are stored in the ion library a total of 1642 proteins were quantified across all samples. On average, 1127 ± 240 proteins were quantified per sample. A maximum of 1560 proteins and a minimum 514 proteins were quantified per sample. After sum normalization, the data set was analyzed regarding the type of distribution. Here, Q-Q-plots were drawn for each sample independently (*cf.* **Appendix 8.3.2**). The plots allowed for the conclusion that the present data does not feature a normal distribution.

As the first step, observed protein abundance was tested for statistical significant changes via the Mann-Whitney u test by comparing the clinical groups (mTBI, Controls and ICB) as used in the discovery study (*cf.* **Chapter 4.3.4**): mTBI vs. Controls and ICB samples; mTBI vs. only Control samples. Within the first group-wise comparison, only seven proteins with a statistical significant ($p < 0.05$) protein abundance change were observed. Respectively, the second group (comparing only mTBI to Controls) revealed 17 protein groups with significant abundance changes. The intersection between these two comparisons featured only three protein groups, namely, P30740 (Leukocyte elastase inhibitor), P06702 (Protein S100-A9) and P12273 (Prolactin-inducible protein).

Table 4.4: List of dual significant protein groups.

No.	Protein Accession	p-value (mTBI vs. rest)	p-value (mTBI vs. CT)
1	P30740 (SERPINB1)	0.010513244	0.020379874
2	P06702 (S100A9)	0.019408045	0.008913398
3	P12273 (PIP)	0.034283299	0.000551833

The direct comparison of protein groups found with a statistically significant abundance change from this analysis and those protein groups found within the discovery cohort (*cf.* **Section 4.3.3**), no accordance was found. None of the previously possible biomarker candidates could be verified in this follow-up and extended cohort. Additionally, in this second study, the list of significantly abundant protein groups is much shorter than in the previous one. Given these results, the list of identified biomarker candidates might then reflect abundance changes introduced through the sample processing process. Any further validation process of the previously identified biomarker candidates appears to be obsolete. Furthermore, any multiple testing correction or false-discovery correction of this newly identified three biomarker candidates in the context of a total 1642 quantified proteins does denominate these candidates as false positive. Hence, it appears that urinary exosomes might not feature exploitable information to discover objectively measurable features to diagnose mTBI.

Nevertheless, proteins of the S100 protein family are discussed in the literature as potential biomarker¹⁸⁴⁻¹⁸⁶. These studies however, describe their findings in the peripheral blood or cerebrospinal fluid (CSF) and correlate their findings to the fact that S100- β monomers are specifically

found in astrocytes and hence that the occurrence of this specific monomer in blood or CSF indicates astrocyte disrupter and in turn also concussion. Here, also a member of the S100 family has been observed with statistical significant ($p < 0.05$) abundance variations between mTBI samples and Controls as well as to ICB samples. It can be hypothesized that the monomers are filtered in the glomeruli (kidney) and hence end up in the urine or in exosomes within the urine. However, S100A9 in particular, has already been described as a key-player in the renal functions¹⁸⁷⁻¹⁸⁹. Hence, these statistically significant protein group abundance changes could rather be correlated to interpersonal renal function alterations than to an actual existent mild traumatic brain injury.

4.3.5 Intracranial-bleed Biomarker Discovery in Urinary Exosomes

Exosomes can perform signaling functions between different tissues as they can be transported via different body fluids to distant parts of the body. Here, besides the initial hypothesis to discover mTBI-related biomarker candidates in urinary exosomes, it is also hypothesized that exosomes originating from traumatic brain injuries are filtered into urine and alter the urinary exosome proteome.

Here, after sum normalization, the protein abundances were also tested for statistical significant abundance changes between, ICB and control with mTBI samples as well as between ICB and control samples alone. Already within the initial sample cohort (21 DIA samples), 31 protein groups were observed with significant abundance change not only within the first comparison set but in both. Nonetheless, the PCA plots generated, based on all quantified protein groups did not indicate any ICB group-based clustering. Additionally, in the extended sample cohort, a total of 80 protein groups demonstrated significant ($p < 0.05$) abundance changes. The contrasting juxtaposition of initial and extended sample cohort revealed an overlap of three protein groups (*cf. Table 4.5*). Here, Q9Y2A7 (Nck-associated protein 1), Q16348 (Solute carrier family 15 member 2) and Q96F07 (Cytoplasmic FMR1-interacting protein 2) were identified in both cohorts with significantly different protein abundance ($p < 0.05$).

Table 4.5: List of protein groups observed as significantly different in both cohorts.

No.	Protein Accession	p-value (1 st Cohort)	p-value (2 nd Cohort)
1	Q9Y2A7 (NCKAP1)	0.037635314	0.025347319
2	Q16348 (SLC15A2)	0.037635314	0.013499121
3	Q96F07 (CYFIP2)	0.047220904	0.031485322

In this context, Q9Y2A7 as well as Q96F07, feature functions that can be related to processes involved in TBIs or secondary brain injuries. The GO annotation revealed that the Nck-associated protein 1 (Q9Y2A7) is associated with inducing apoptotic processes while modulating actin reorganization processes via the WAVE protein complex. Additionally, a possible relation to Alzheimer's disease (AD) has also already been described¹⁹⁰⁻¹⁹². In this regard, AD is recognized as a neurodegenerative disease, which involves neuronal cell death due to various involved disease mechanisms. Nevertheless, brain injury can be counted as secondary injury in this context

of AD which consequently leads to the assumption that NCKAP1 occurrence in urinary exosomes relates to an existing brain bleed in patients. Besides the involvement of NCKAP1 in neurological apoptosis process, the GO annotation of Cytoplasmic FMR1-interacting protein 2 (Q96F07) revealed that this protein is involved in the T-cell adhesion and p53/TP53-dependent induction of apoptosis¹⁹³⁻¹⁹⁵. Additionally, due to its high similarity to Q5SQX6 (CYFP2_MOUSE), the identical protein in mice, it is deviated that this protein is highly expressed in the perinuclear region and enriched in synaptosomes¹⁹⁶.

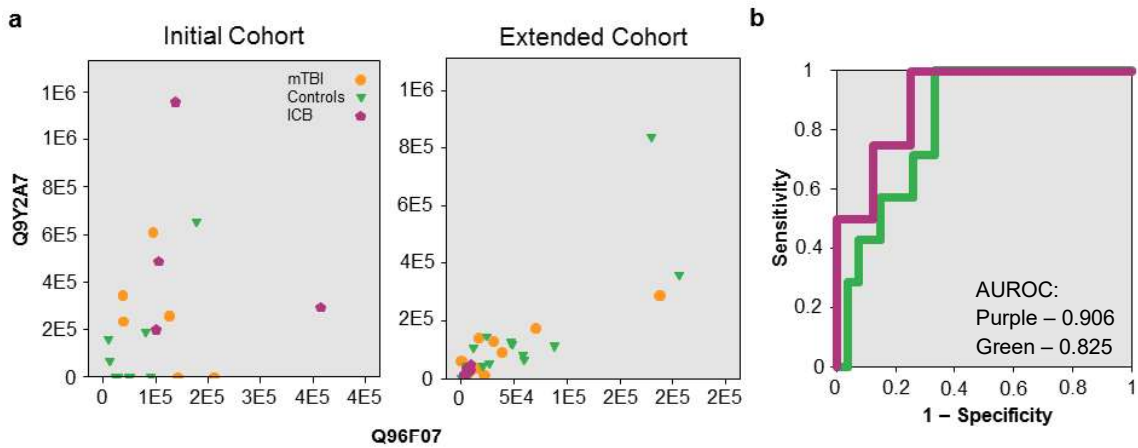


Figure 4.5: Statistical analysis of Q9Y2A7 and Q96F07. a) 2D-scatterplots of Q9Y2A7 and Q96F07 for initial and extended sample cohort. Scatter-plot features a spread while for the extended cohort, ICB samples cluster at the origin of the scatter plot. b) ROC curve for the linear regression of Q9Y2A7 and Q96F07 based on either the initial discovery cohort (purple) and the extended sample cohort for qualification (green). AUROC for both curves is > 0.8.

Further, the diagnosing applicability of these proteins in the context of intracranial bleed was investigated. At first, the AUROC for each protein alone was identified. Here, each protein featured an AUROC of > 0.7 in both investigated cohorts. In addition to the investigation for each protein alone, the linear combination of the two neuronal proteins was also tested for biomarker applicability. Corresponding AUROC values are given in **Table 4.6**.

Table 4.6: List of calculated AUROC values. ROC curves for the linear combination can be found in **Figure 4.5**.

Accessions	Discovery Cohort	Qualification Cohort
Q9Y2A7	0.844	0.778
Q16348	0.844	0.807
Q96F07	0.828	0.815
Q9Y2A7 and Q96F07	0.906	0.825

For both cohorts, the linear combination of the neurological proteins (Q9Y2A7 and Q96F07) the AUROC was enhanced compared to their single counterpart. Nonetheless, perfect biomarker conditions (AUROC of 1) were not reached. For both cohorts the linear combination was visualized via 2D scatterplot (*cf.* **Figure 4.5**). For the initial cohort, ICB samples are spread across higher protein values, while the control and mTBI samples are found closer to the origin of the plot. In contrast, ICB samples from the qualification cohort cluster closer to the origin, while control

and mTBI samples spread towards higher abundance levels. Based on the associated biological process (BP) and the sub-cellular location of the protein itself, a higher protein abundance appears logical. Nonetheless, the effect causing the statistical significance is inverted between the initial and qualification cohort. Besides the fact that these proteins appear with statistically significant protein abundance changes and their statistical usability for objectively diagnosing intracranial bleeds, the occurrence of effect inversion definitely compromises the applicability of these proteins as clinical biomarker for diagnosing intracranial bleed via urinary exosomes.

4.3.5.1 Analysis for involvement in disease or functional processes

Finally, the association of protein groups with significant abundant changes to certain disease or function specific pathways was investigated. The scope of this investigation was the identification of processes which are specifically involved in neuronal damage, neurodegeneration or any other neurological impairments, which give rise to the presence of an intracranial bleed. Nonetheless, for a biomarker it is not theoretically an essential event to be correlated to a disease-specific pathway¹⁹⁷, however, in the case of a link being found, the actual marker might be additionally exploited as a drug target. Here, the association of proteins to disease and function-specific processes help to identify proteins or protein groups which are involved in the actual neurological injury process due to intracranial bleed.

The analysis of the protein list originating from the initial sample cohort did feature certain involvements in neurological disease pathways. Here, the key protein was SOD1 (P00441; Superoxide dismutase [Cu-Zn]). This protein is described as an enzyme specifically destroying radicals which are normally produced within the cells and which are toxic to biological systems. Besides SOD1 also the aforementioned NCKAP1 (Q9Y2A7 has been associated with the differentiation of cortical neurons. Otherwise, no further involvements in neurological pathological pathways or general neurological process was found.

In contrast, the extended sample cohort features 16 protein groups which were associated with neurological impairments. Many noted impairments (*cf.* **Appendix 8.3.3**) refer to abnormal morphology of neuronal components, as e.g. abnormal morphology of brain. Here, the top three associations with neuronal dysfunctions are abnormal morphology of cerebellum, abnormal morphology of fourth cerebral ventricle and abnormal morphology of brainstem with p-values of $> 7E-4$. Surprisingly, none of the aforementioned and further investigated biomarker candidate molecules was assigned to any neurological mechanism.

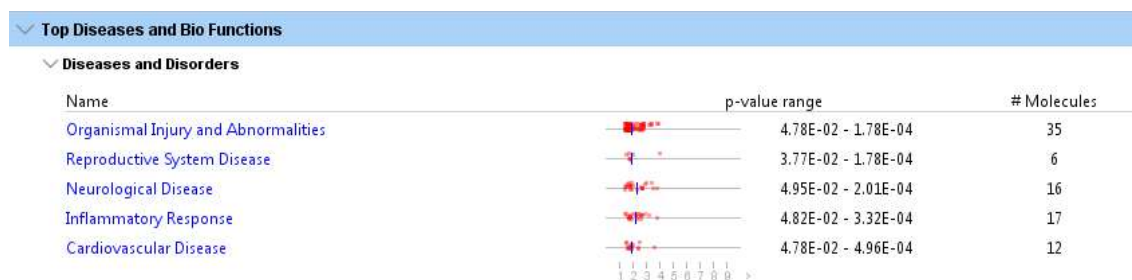


Figure 4.6: Screenshot of Ingenuity® IPA® summary page. On the summary page, “Neurological Diseases” already stands out as a top three match of the uploaded protein entries. The corresponding p-values for identified processes lies within ~ 5 to $2E-4$. A total of 16 protein groups ID was able to be assigned to this functional and disease group.

4.4 Perspective

Exosomes represent a highly valuable research target for the discovery of novel biomarkers for certain diseases. The experiments conducted have shown that processes from even very distant body parts can be verified in urinary exosomes. Furthermore, exosomes allow insights into cellular processes and give rise to certain signaling attempts.

Here, biomarker candidates for mild traumatic brain injury were identified in the first sample cohort, but could not be validated in the admitted qualification cohort. On one hand this can be explained by either the rather small cohort size and hence interpersonal difference can be picked up as statistical significance. On the other hand, the Tamm-Horsfall protein (Uromodulin; P07911) imposes a giant adversary for the exosome extraction¹⁷⁵. This highly abundant protein in urine can co-complex certain urinary proteins which consequently co-sediments with the exosomes during ultra-centrifugation. However, even extensive washing of the exosome pellets might not abolish these uromodulin-clusters. Furthermore, exosomes might even be enclosed in these clusters¹⁷⁵. Here, the expansion of the sample cohort to specifically discover biomarker candidates for mTBI appear to be necessary. Furthermore, the data has shown that the proteome within exosomes might not be altered due to sustain a concussion.

Besides the attempt to discover mTBI-related biomarker candidates in urinary exosomes, this study setup also allowed the investigation for biomarker candidates for intracranial bleed. Here, a panel of biomarker candidates was discovered that allowed for objectively diagnosing the presence of an intracranial bleed. Furthermore, the candidate molecules were observed within two independent sample cohorts. Nonetheless, further evaluation via other methodology (e.g. Western Blot or ELISA) is required before any advancements can be undertaken. Furthermore, many associations of proteins with neuronal pathological processes were found. However, a biomarker candidate could not be deviated as these proteins feature rather poor clinical satisfaction with regard to sensitivity and/or specificity. Nonetheless, intracranial bleeds can be divided into four different types: intracerebral (parenchymal; within the brain matter), epidural, subdural and sub-arachnoid¹⁹⁸⁻²⁰⁰. Yet, it is not known to what extent the certain types of intracranial bleed can further alter the proteome network. Hence, further research has to be conducted to fully understand the impact of intracranial bleed type and proteome composition. Ideally, the exosome proteome can help to stratify amongst the mentioned types of ICB. Both sample cohorts for this investigation were lacking any ICB classification, leaving meta-information in a binary state (bleed or not).

Over the few past years, exosomes have receive an increasing amount of attention as a research target, especially regarding the discovery of novel biomarkers. Furthermore, the clinical applicability of biomarkers found in exosomes has been extensively discussed and is expected, especially since technological advancements allow for reliable and cost-effective exosome extraction^{201, 202}. Nonetheless, it appears that the clinical applicability of exosomal mTBI biomarker is compromised, especially by the time factor. Clinicians or healthcare provider need almost instant results to adjust their therapeutic approach. Hence, diagnostic exosomal mTBI biomarkers will provide no benefit. Additionally, intracranial bleeds need to be ruled-in or ruled-out immediately,

which can be easily achieved by imaging technologies like CT or MRI. However, urinary exosomes, due to their non-invasive recovery have gained a special status and can provide insightful information as well as provide valuable biomarkers. Lastly, exosomal biomarker in the context of time-sensitive injuries as mTBI or intracranial bleed, appear not to be an obvious research area.

4.5 Acknowledgments and Contributions

I want to thank Dr. Nerea Cuevas Polo for her preliminary evaluation of strategies for successful exosome extraction. Furthermore, I want to thank Dr. Judith Steen and Dr. Hanno Steen for insightful discussions and scientific support. Last but not least, I want to thank Dr. Rebekah Mannix and Dr. Mark Berry for the establishment of the urinary sample cohort and all various research coordinators for the collection of urine samples in the Emergency Medicine Department at Boston Children's Hospital (Boston, MA, USA).

My personal contributions to this project account for: exosomal preparation methodology optimizations, carrying out the described experiments, analyzing the samples with mass spectrometry, data analysis and figure generation.

5. Terminomal Biomarker Discovery for mTBI

5.1 Introduction

The bottom-up proteomics approach describes the specific digestion of proteins or even protein mixtures with specific proteases. Resulting peptides hence feature specific N- or C-termini which are specifically used to identify proteolytic peptides and thus to identify proteins based on the peptide sequence snippet. In this context, generic and spontaneous protein degradation is mainly avoided by storing proteins at sub-zero temperatures and especially working with protein solutions at low temperatures (e.g. 4°C). Here, it is desirable to avoid any falsely introduced protein degradation or cleavage through external proteases. In this context, proteomic search algorithms can be specifically set by the user to ignore any peptide which does not feature designated cleavage sites originating from the digestive enzyme used.

Proteins, which circulate throughout the body, are steadily exposed to temperatures, proteases, pH conditions or pathogens like bacteria or viruses which can cause spontaneous protein degradation. However, it is widely known that pathogens can specifically degrade, or i.e. cleave host related immune proteins by secreted proteases²⁰³⁻²⁰⁸ to evade the hosts' immune defense. Additionally, endogenous occurring proteases (e.g. metalloproteases) can also cleave useful proteins just by recognizing its characteristic cleavage pattern. Nonetheless, uncontrolled proteolysis reactions are sometimes part of certain diseases such as Synucleinopathies^{209, 210} or cancer²¹⁰⁻²¹². Urine is the by-product of filtering plasma through the glomeruli in the kidneys^{213, 214}. Here, unwanted and unnecessary molecules such as metabolites, proteins, etc., are expelled from the blood and later excreted via urination. Urine is not constantly excreted but stored for a certain time in the bladder. Under certain circumstances, urine is stored for several hours in the bladder, where it is technically "incubated" at 37°C, body temperature. Practically, this imposes ideal working conditions for proteases to cleave proteins. In this context it was hypothesized that due to either disease-specific or due to endogenous protein degradation, peptides with altered N- or C-termini, apart from termini established through the employed digestive enzyme for sample preparation, can be identified and further quantified²¹⁵. Based on the scope of research, semi-specific peptides and their altered termini, the terminology "Terminomics" was deviated.

In this study, semi-tryptic peptides are identified and then quantified to discover disease-specific proteolytic events which in turn could provide diagnostic and/or other biomarker-like indications.

5.2 Experimental Procedures

5.2.1 Identification of semi-specific peptides

For this study, no extra urine samples were acquired. The basis for this analysis forms the qualitative (DDA) and quantitative (DIA) data from the aforementioned mTBI urine biomarker qualification study (*cf.* **Section 3.3.3**).

After the quantification process of the 50 DIA files was completed (*cf.* **Section 3.2.4.2**), the list of quantified peptides was specifically filtered for peptides, which demonstrated either a non-tryptic N- or C- terminus. Hereto, an in-house PHP script was employed, which directly compared the identified peptide sequence with the corresponding protein sequence.

In short, the quantitative output report features one line for each quantified peptide, or precursor. Here, the assigned protein group or protein accession to which the precursor has been assigned, is also given. Hence, after parsing the list of peptides into an associated array, the corresponding protein sequence, based on the given protein accession by the search algorithm, was directly loaded from the UniProt database by using their provided application programming interface (API). Consequently, the peptide sequence was matched to the protein sequence. After obtaining the position of the first peptide amino acid within the protein sequence, the N-terminus cleavage site was analyzed for trypsin specificity by testing the (n-1) position for presence of R (Argenine) or K (Lysine). Consequently, the C-terminus of the peptide was tested for trypsin specificity by again testing for presence of K or R. Based on this classification, either the peptide was deleted out of the associated array (fully tryptic) or a classification string was added to the peptide array. Finally, the associated array was written out as tap-separated text file for further analysis.

5.2.2 Peptide-based Quantification

The filtered peptide file was further modified, allowing for statistical group-wise comparison. Here, each peptide was re-assigned to its parent protein. In short, after loading the filtered peptide output file was loaded into EXCEL (Microsoft, USA), the data was condensed using the built-in “Pivot Table” options. Here, peptides could easily be remapped to their parent protein, allowing for peptide quantification within the assigned protein group.

5.2.3 Statistical Analysis

Quantitative values were statistically analyzed with either SPSS v23 (IBM) or EXCEL with “Real Statistics Resource Pack” plugin. Further statistical analysis was performed with the proteomics data analysis tool: InfernoRDN (v.1.1.6044) ¹⁷⁴ from <http://omics.pnl.gov> (U.S. Department of Energy).

5.3 Results

In the dataset a total of 787 non-specific, i.e. semi-specific peptides were identified. Peptides associated with general laboratory proteins were deleted from the target list prior to peptide analysis. Out of the proteins identified, 227 peptides related to peptides featuring a non-tryptic cleavage at the C-terminus and 560 peptides with a semi-specific N-terminus. During the filtration process for semi-tryptic peptides, N-terminal or C-terminal peptides were also specifically excluded from consideration by a special implemented screening algorithm. Consequent peptide quantitation was conducted by summing the “MS2PeakArea” for each peptide. Consequently, the data distribution was tested via Q-Q-plots. Here, normal distribution was ruled out, as no diagonal line was observed in any sample-based plots (graphs not displayed). Hence, for any further analysis, non-parametric conditions were presumed.

5.3.1 mTBI-based analysis

At first, peptide quantities originating from the mTBI group were statistically compared to those originating from Controls and ICB group. For this first comparison, Controls as well as ICB samples were both considered control samples. In this first analysis a total of 39 peptides were identified with a p-value < 0.05. In this discovery step, p-values were acknowledged as is, and were not further corrected by multiple-testing (e.g. Bonferroni or Benjamini-Hochberg). Secondly, ICB samples were removed from the “Control” group and hence, mTBI samples were compared only to controls. Within this comparison a total of 32 peptides was identified with a p-value of less than 0.05. Nonetheless, no peptide featured a significance lower than $p = 0.001$.

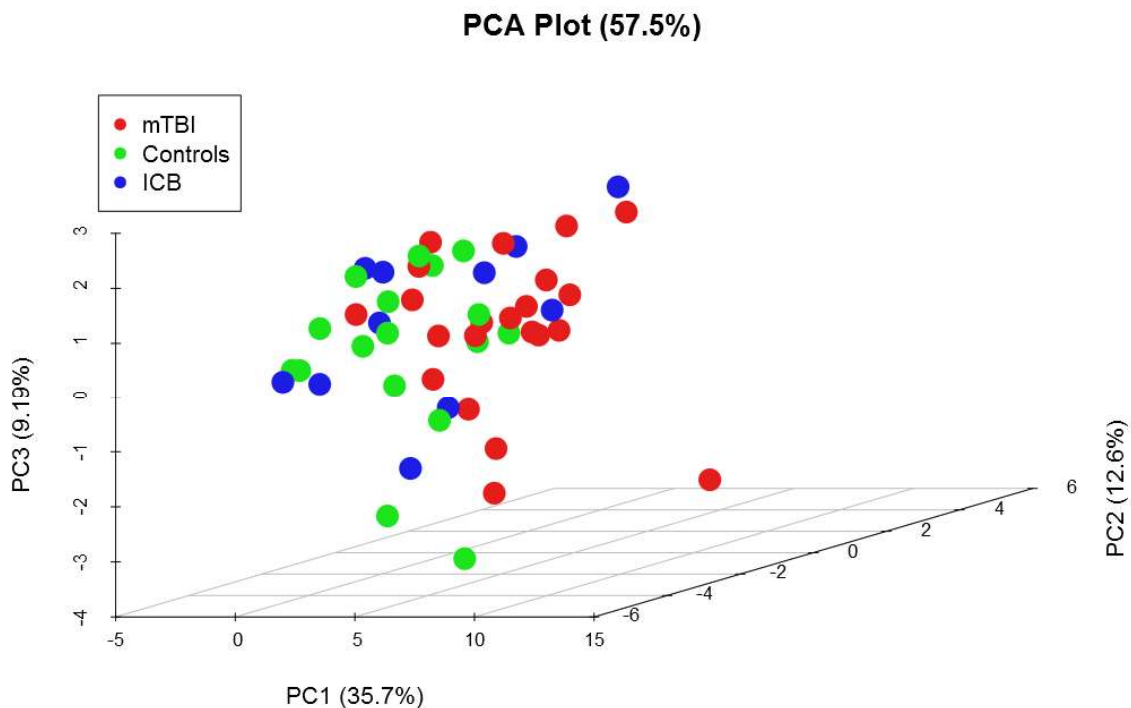


Figure 5.1: 3D PCA plot for significant observed terminomal peptides. PCA loadings based on significantly observed quantity variation of peptides with semi-specific cleavage. The plot accounts for 57.5% of the dataset variation, whereas PC1 accounts for 35.7%, PC2 for 12.6% and PC3 for 9.2% variation. No clear group wise clustering observable.

As consequent investigation, the intersection between the first statistical group comparison and the second was estimated. Here, 22 peptides (*cf.* **Appendix - Table 8.4.1**) showed significance ($p < 0.05$) in both comparisons. The peptides that were filtered here were entered into a PCA (*cf.* **Figure 5.1**) to identify clustering tendencies in the peptide quantities. The corresponding three-dimensional PCA plots (PC1, PC2 and PC3) accounts for a total of 57.5% of the data variation. Additionally, PC1 accounts for 35.7%, PC2 for 12.6% and PC3 for 9.2%. Nonetheless, no clear or obvious group-based clustering and separation was observed in the PCA plot.

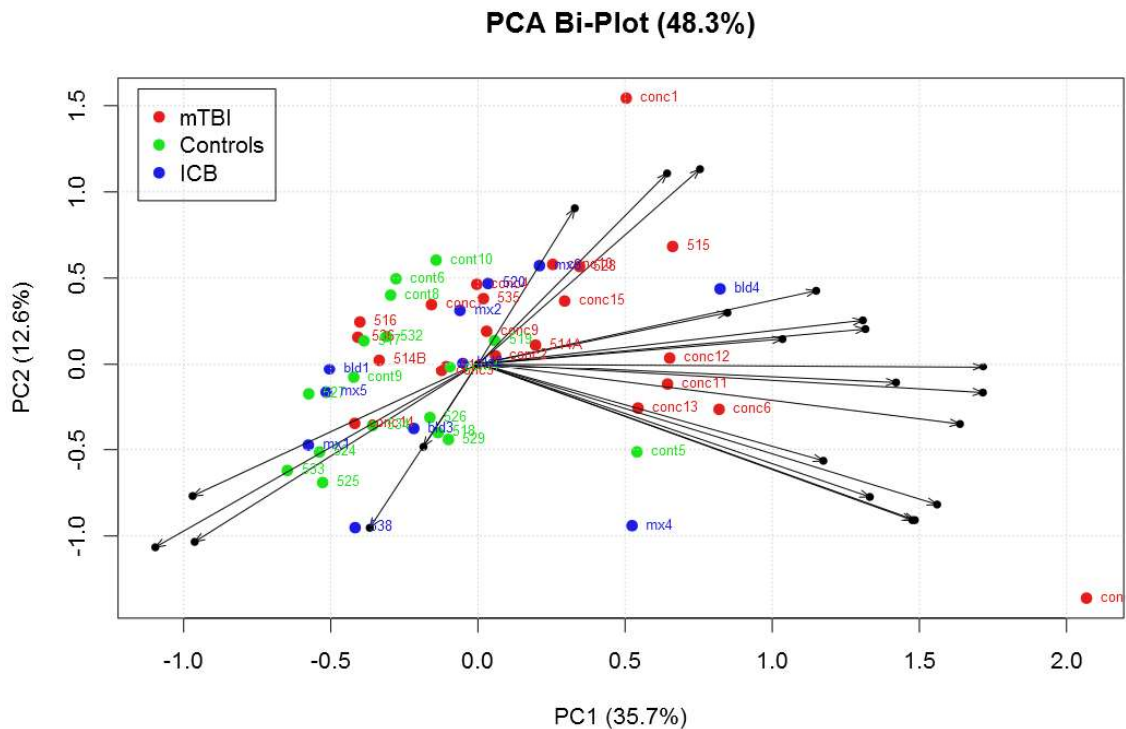


Figure 5.2: Supervised PCA Bi-Plot for mTBI biomarker discovery. All significantly observed peptides entered into PC analysis. There is no peptide specific separation of clinical groups. While Controls (green) rather spread along PC 2, mTBI (red) samples spread across PC 1. PCA blot accounts for a total of 48.3 % of the dataset variance.

Furthermore, the identified and significantly different peptides that were observed, were traced back to their parent protein. Protein accessions were used to further identify pathways or biological processes that could have been compromised or specifically regulated by protein degradation. Out of the 22 previously identified peptides, 18 proteins and their corresponding accessions where deviated. Here, amongst the results of the Ingenuity® IPA® pathway analysis, pathways listed with the most significant p-values are: LXR/RXR activation ($1.4E-8$), FXR/RXR activation ($1.7E-8$) and Acute Phase Response Signaling ($4.9E-5$). Furthermore, “Immune Response” was listed as possible “Disease or Disorder” with a p-value range of $4.41E-2$ to $6.35E-5$. However, no specific pathway for neurological processes was discovered amongst the submitted protein accessions.

5.3.2 Intracranial bleed focused analysis

Besides the analysis for the discovery of mTBI based processed, here an analysis for ICB-based processes was also performed. Initially, the peptide quantities of ICB sample were statistically compared to mTBI as well as control samples via Mann-Whitney u test. In this first comparison, mTBI and controls were jointly considered one group. Here, a total of 39 peptides were identified demonstrating a significance level of $p < 0.05$. For discovery purposes, p-values were not yet corrected for multiple testing. Consequently, ICB-related peptide quantities were only compared to control samples. This comparison revealed 40 peptides with $p < 0.05$. No additional multiple testing correction was applied here for discovery purposes. Consequently, the intersection of both comparisons was assessed. In total, 20 peptides were observed with significant peptide quantity changes in both non-parametric group-wise comparisons (*cf. Appendix - Table 8.4.2*).

The aforementioned 20 peptides were entered into a PC analysis to estimate their diagnostic value. Here, the first principal component (PC1) accounts for a total variance of 39.8%, while PC2 accounts for 12.7%. The third principal component (PC3) describes another 9% of the dataset variations. Hence, a 2D-PCA plot (PC1 & PC2; *cf. Appendix 8.4*) displays a total of 52.6% of the dataset variation, while the combined 3D-PCA plot (PC1, PC2 & PC3) accounts for 61.5% (*cf. Figure 5.3*) of the variance. In both plots, a tendency of ICB sample clustering can be observed.

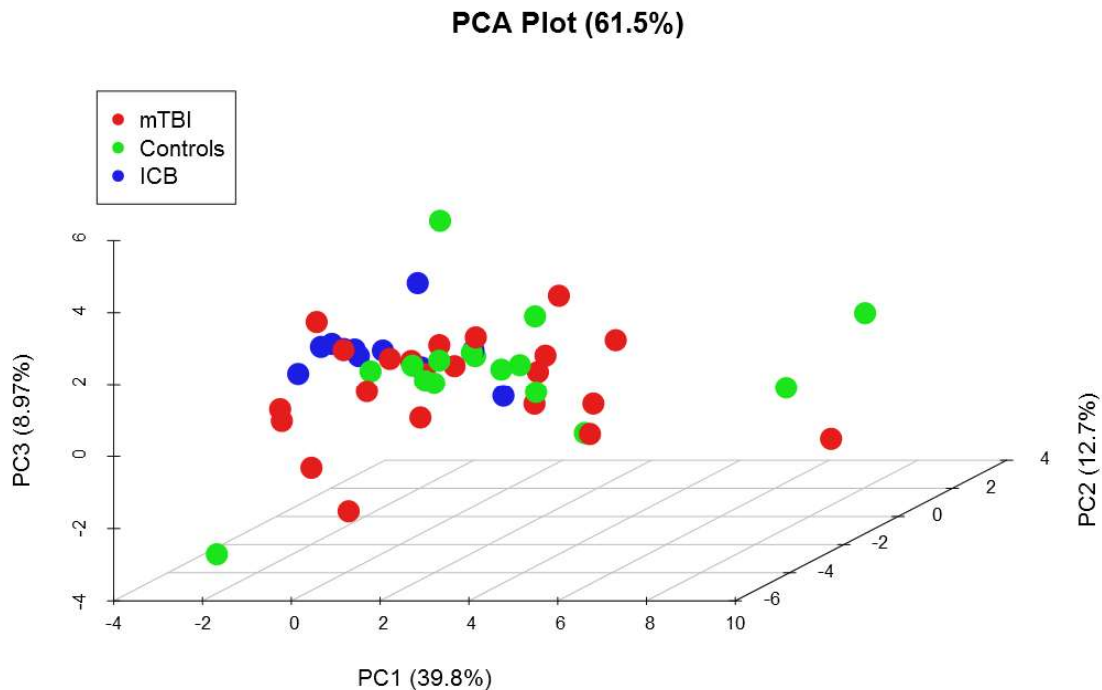


Figure 5.3: 3D PCA plot for the discovery of ICB specific clustering. PCA plots for the top three principal components: PC1 representing 39.8 % variance, PC2 accounting for 12.7 % variance and PC3 featuring ~9% dataset variance. ICB samples tend to cluster in the top left corner while one ICB sample is rather centric. PCA plots demonstrates a total of 61.5% of the dataset variance.

All but one of the samples (bld2) are clustered to the top left corner of the plot. Nonetheless, an absolutely clear separation of ICB samples from all others is not achieved. The corresponding PCA Bi-plot (*cf. Figure 5.4*) reveals that only a peptide with the sequence “GDELLRFSN”

causes the clustering of ICB samples into the top left corner: the corresponding AUROC numbers 0.718 (ROC plot not shown). Furthermore, based on the peptide sequence, it was reciprocal associated with the AMBP protein (P02760). This protein consists of 352 amino acids and features a total molecular weight of 40 kDa, as well as the protein exhibiting different interaction sides. Binding sites occur within the head of the protein (amino acid #53 to #149) chromophore. Towards the end of the polypeptide chain, this particular protein features a trypsin-specific inhibitory functional side. In this context, the peptide of interest was originating from a C-terminal cleavage of “Proline” at position #340 in the amino acid chain. The trypsin inhibitor region however, is indicated to range from #297 to #298.

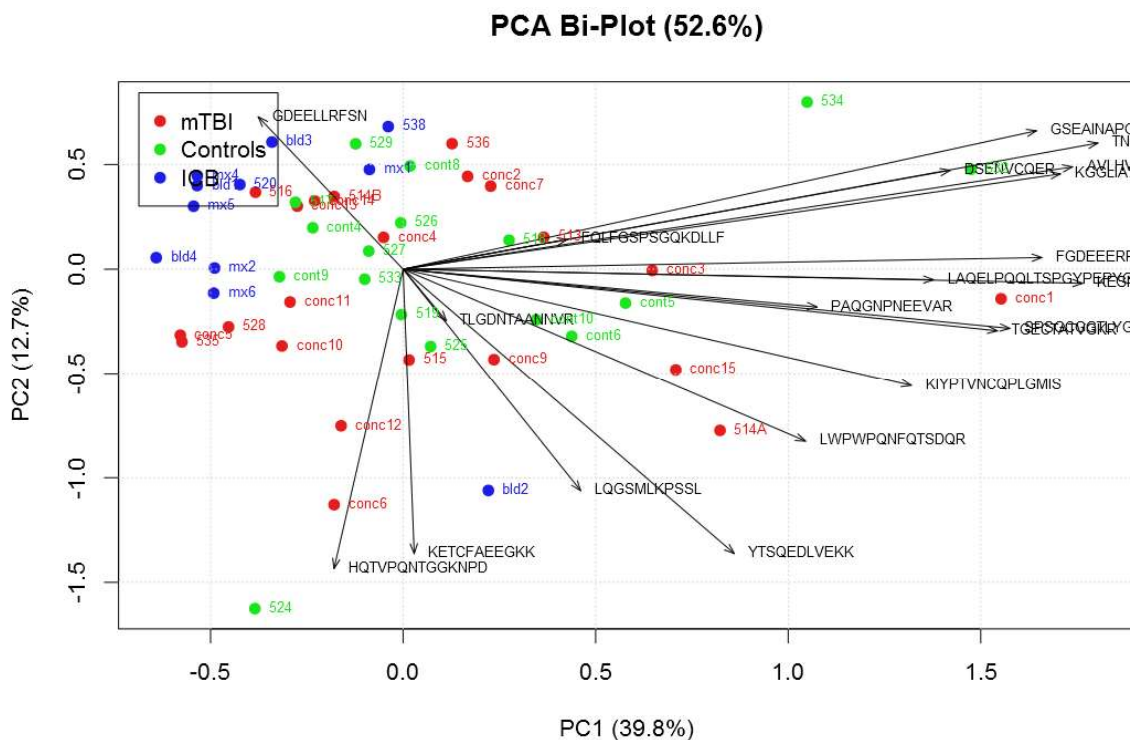


Figure 5.4: PCA Bi-Plot of significantly observed terminomal peptides in ICB. The plot is based on two principal components: PC1 and PC2. Components account for a total of 52.8% dataset variance. Components account for either 39.8% (PC1) or 12.7% (PC2) of the dataset variance. Single peptide “GDELLRFSN” causes a cluster of ICB samples towards upper left corner.

Apart from the PCA analysis, peptide information was also entered into a pathway-centered analysis. Here, the statistically significant peptides originated from 17 proteins (*cf. Appendix -Table 8.4.2*). Consequently, the protein accessions were fed into the Ingenuity® IPA® pathway analysis algorithm. Here, a total of six proteins were non-exclusively associated with neurological disorders. Via KNG1 (P01042), a particular link to “damage of blood-brain barrier” with a p-value of $5.55E-3$ was found. Furthermore, Kininogen-1 was represented in the dataset with 13 semi-specific tryptic peptides. However, only one peptide demonstrated significance across both applied comparisons.

5.4 Discussion

The proteome is always tightly controlled by processes in the body. Post-translational modifications add extra layers of control to this system. Nonetheless, proteolysis probably imposes the most effective mode of control²¹⁵. Proteomic studies mainly focus on the identification of tryptic peptides, which should ideally not feature any cleavage alterations. Nonetheless, the search for semi-specific peptides and their consequent identification would cause a higher peptide coverage, which additionally enables the identification of more proteins in complex samples²¹⁶. However, computational processing times are massively increased due to the enhanced possible peptide precursors. Here, the proteolytic proteome control process is exploited to possibly deviate disease-related diagnostic information from proteins that feature specific cleavages from endogenous proteases other than the uniformly used trypsin during sample preparation for mass spectrometric analysis.

In this study, various semi-specific peptides have been identified and in this context, the question remains of whether the detected non-tryptic cleavage was caused by spontaneous protein degradation or just randomly, or whether it actually originates from directed protein cleavage through endogenous proteases. Nonetheless, the acquisition of a mass spectrometric signal (actual peak) of a particular precursor is based on a high number of identical ion species which have to be present in the sample. Based on this fact that semi-specific peptides which were acquired with a signal higher than the noise level are only based on spontaneous protein degradation as too many identical ion species are present to entitle the process as random event, can be excluded.

The abundance comparison of semi-tryptic peptides between mTBI samples and controls as well as ICB samples lead to a list of possible diagnostic target molecules. Nonetheless, a valuable diagnostic performance could not be deviated. Whether or not the PCA analysis demonstrated diagnosing tendencies, the reciprocal pathway analysis did not reveal any neurological related pathways or disease processes.

In contrast, within the ICB samples it was observed that these samples collectively share a semi-specific peptide that inherited classification tendencies. The analysis of this peptide revealed a N-term based cleavage after (c-terminally) Prolin. Trypsin per se, features tendencies to also recognize Prolin as a cleavage site. If in this case, the cleavage was caused by trypsin, the prolin amino acid would be part of the observed peptide, as trypsin cleaves N-term specific on prolin²¹⁷. However, also shown here in the PCA plot, no clear separation could be achieved. In this context, the satisfactory clinical expectations were not achieved by simply basing the diagnosis on this single peptide. Additionally, the deviation of a clinical valuable diagnosis based on a single semi-specific peptide appeared too vague, which lead to the consequent refusal of this lead.

Thus far, the analysis of semi-specific peptides in the context of clinically usable diagnosing information has not revealed any useful insights. Nonetheless, this type of investigation might contain insightful information in the context of diseases other than neurological disorders.

5.5 Acknowledgments and Contributions

I want to thank Dr. Hanno Steen and Dr. Jan Muntel for their input and assistance during establishing this research idea. Furthermore, I want to thank Dr. Shaojun Tang and Dr. Kevin Broadbelt for insightful discussions and scientific support. Last but not least, I want to thank Dr. Rebekah Mannix and Dr. Mark Berry for the establishment of the urinary sample cohort and all various research coordinators for the collection of urine samples in the Emergency Medicine Department at Boston Children's Hospital (Boston, MA, USA).

My personal contributions to this project account for: input to the initial research idea, carrying out the described experiments, analyzing the samples by mass spectrometry based proteomics, data analysis, programming the in-house PHP script and figure generation.

6. CSF-based Biomarker Discovery for NDs

6.1 Introduction

Every year the mean age of the world population increases which in turn, also influences the incidences of age-related disorders ²¹⁸. Amongst these disorders, neurodegenerative diseases have drawn lots of attention as their progress is mostly irreversible, they feature a lack of effective treatment options as well as being heavily associated with social and economic burdens ²¹⁹. Nonetheless, neurodegenerative diseases feature different clinical patterns, which can however overlap to a great extent. Common symptoms of neurodegenerative diseases can be described as: paralysis, tremors, muscle weakness, poor coordination, dementia, loss of sensation, confusion or altered levels of consciousness.

On the cellular level, neurodegenerative diseases (ND) are described as the progressive loss of neuronal cells in the brain. With the increased disappearance brain cells, the aforementioned symptoms also worsen or further symptoms can manifest in the clinical picture. Furthermore, on the molecular level, NDs are mainly described as proteopathies, as the diseases are caused through misfolded proteins which cause the consequent cell death. However, misfolded proteins are not the only cause of the onset of neurodegenerative disorders. On the molecular level, different reasons, such as genetic impairments ^{89, 220}, mitochondrial dysfunction ^{221, 222} or activation of programmed cell death ²²³ are widely discussed. Even though causes are known, diseases remain incurable once they have been diagnosed. Hence, the major goal of research lies within the discovery of methodological approaches or biomarkers, that in turn allow for the early detection of neurodegenerative diseases as well as the presymptomatic detection of neuronal dysfunction ²²⁴.

The brain features a certain fluidic layer between the skull and the cortex. This fluidic layer is furthermore defined as cerebrospinal fluid (CSF) which hence provides a mechanical and immunological protection for the brain. Additionally, the brain is fully submerged in CSF, which also reaches further down into the spinal cord ²²⁵. Besides the protective function of CSF for the brain, CSF also adopts the function of the waste management of the brain tissue. Waste products or toxic molecules can hence easily be removed by the regular flow of CSF in the skull. Here, CSF can demonstrate a certain “sink” function, which allows the further transport of waste into the bloodstream and hence beyond the blood-brain barrier. Due to its close proximity to the central nervous system (CNS) as well as to affected brain regions by degenerative diseases, CSF imposes a valuable body fluid for researchers in the context of neurodegenerative diseases. In particular, CSF can reflect metabolic processes as well as the biochemical state of the brain.

In this study the CSF of patients suffering from four different neurodegenerative disorders, namely corticobasal degeneration (CBD), Parkinson’s disease (PD), Alzheimer’s disease (AD) and progressive supranuclear palsy (PSP), were analyzed. Patients displaying “healthy” conditions are enrolled as control samples. Consequently, samples were analyzed with the aim of the discovery of diagnostic as well as stratification biomarkers.

6.2 Experimental Procedure

6.2.1 CSF sample collection

Samples were collected at the University of California San Francisco (UCSF), USA. After collection, samples were entered into the specimen bank at the UCSF Memory and Aging Center. The sample cohort was retrieved as part of a collaborative project. Corresponding sample tubes contained about 1 ml of neat cerebrospinal fluid (CSF) each. Signed consent was obtained before sample collection. Cohort acquisition was under the supervision of the Institutional Review Board (IRB) at UCSF.

6.2.2 MStern Blot

The CSF sample was thawed on ice and the neat sample (12 μ l, i.e. \sim 12 μ g of protein) was diluted in 138 μ l 50 mM ABC. Hereafter, the diluted sample (150 μ l sample volume) was added to a mixture of 150 μ g urea and 30 μ l dithiothreitol (DTT) (100 mM in 1 M Tris/HCl pH 8.5). The resulting solution was incubated for 20 min at 27°C and 1100 rpm in a thermo mixer. Reduced cysteine side chains were alkylated with 50 mM IAA (final concentration) and incubated for 20 min in the dark at 27°C and 750 rpm.

The hydrophobic PVDF membrane in a 96-well plate format (MSIPS4510, Millipore) was pre-wetted with 150 μ l of 70% ethanol and equilibrated with 300 μ l urea supernatant (\sim 8.3 M urea). These and all subsequent liquid transfers were carried out using a fitted 96-well micro-plate vacuum manifold (MAVM0960R, Millipore).

Each sample was drawn three times through the PVDF membrane by applying a vacuum. After protein adsorption onto the membrane, it was washed twice with 50 mM ABC. Protein digestion was performed with sequencing grade trypsin (V5111, Promega) at a nominal enzyme to substrate ratio of 1:15. To this end, 100 μ l digestion buffer (5% acetonitrile (ACN; v/v), 50 mM ABC and trypsin) were added to each well.

After incubation for 2 hours at 37°C in a humidified incubator, the remaining digestion buffer was evacuated. Resulting peptides were eluted twice with 150 μ l of 40% ACN (v/v)/0.1% (v/v) formic acid (FA) each. Upon pooling, the peptide solutions were dried in a vacuum concentrator. Lyophilized samples were stored at -20°C for further analysis.

6.2.3 LC-MS/MS analysis

6.2.3.1 Ion library generation for data-independent acquisition

Dried peptide mixtures were reconstituted in 30 μ l sample buffer (5% ACN / 5% TFA) and needed calibration peptides (HRM Calibration Kit; Biognosys, Switzerland) were spiked-in with a 1:30 ratio (1 μ l peptides per sample). Tryptic digests (\sim 1.2 μ g; 3 μ l sample) were directly injected onto a PicoChip (NewObjective, ReproSil-Pur C18 3 μ m, 75 μ m x 105 mm) and peptides were consequently separated by a linear gradient from 93 % buffer A (0.2 % FA in HPLC water; buffer B: 0.2 % FA in ACN) to 70 % within 120 min. Analytes were ionized and sprayed into the mass spec-

trometer (Q Exactive, Thermo Fisher) by applying 2.2 kV to the spray emitter. The mass spectrometer was operated in data-dependent TOP10 mode with the following settings: mass range 375 – 1400 Th; resolution for MS1 scan 70 000 @ 200 Th; lock mass: 445.120025 Th; resolution for MS2 scan 17 500 @ 200 Th; max. injection time (IT)120 ms; isolation width 1.6 Th; Normalized Collision Energy (NCE) 27; underfill ratio 2.0 %; charge state exclusion: unassigned, 1, >6; dynamic exclusion 30 s.

6.2.3.2 Quantitative data acquisition (DIA)

The same samples as in **Section 6.2.3.1** were used for quantitative analysis. Tryptic peptide mixtures (~1.2 µg) were directly injected onto a PicoChip (NewObjective, ReproSil-Pur C18 3 µm, 75 µm x 105 mm) and peptides were consequently separated by a linear gradient from 93 % buffer A (0.2 % FA in HPLC water; buffer B: 0.2 % FA in ACN) to 70 % within 120 min. Analytes were ionized and sprayed into the mass spectrometer (Q Exactive, Thermo Fisher) by applying 2.2 kV to the spray emitter. Each DIA cycle contained one full scan (MS1) and 17 consecutive target MS² (t-MS²) scans covering a mass range of 400 – 1000 Th. The full scan resolution was set to 35,000 @ 200 Th; AGC target –3e6, maximal IT 120 ms; mass range 400 – 1,000 Th; followed by DIA scans with resolution 35,000 @ 200 Th; variable window size was applied used; NCE 27; AGC target value 1e6, max. IT 120 ms.

Table 6.1: Variable extraction windows used in DIA experiments.

Window No.	Window size
1	26 Da @ 413 m/z
2 to 8	27 Da @ 437.5, 462.5, 487.5, 512.5, 537.5, 562.5 and 587.5 m/z
9 to 12	32 Da @ 615, 645, 675 and 705 m/z
13 to 16	52 Da @ 745, 795, 845 and 895 m/z
17	81 Da @ 959 m/z

6.2.4 DIA data analysis

For ion library generation, all corresponding DDA analysis files (*.raw) were loaded into MaxQuant¹¹⁷ (v.1.5.2.8) and consequently searched against a reference protein sequence database, here, the human UniProtKB database (Homo sapiens, ~20,300 sequences, version 06-2014) was used with commonly occurring laboratory contamination protein sequences (cRAP, version 2012.01.01) added. Variable Modifications were set to: Acetyl(Protein N-Term), Oxidation(M), Deamidation(NQ) and Phospho(STY). For fixed modifications only Carbamidomethyl(C) was used. The digestion mode was set to “Specific” and the digesting enzyme to “Trypsin”. Otherwise default settings (by MaxQuant) were used for the analysis: trypsin with up to two missed cleavages; mass tolerance for the first search: 20 ppm; main search: 4.5 ppm.

DIA files (*.raw), which belong to the analysis of the larger cohort for biomarker qualification, were processed and analyzed in Spectronaut¹⁴⁷ (v.7.0.8065; Biognosys, Switzerland). At first, the search results from the aforementioned MaxQuant search of the corresponding DDA acquisition

were loaded into Spectronaut to generate an ion library, which in turn was saved within the program's data-structure. The data ion library import was performed via the default settings proposed by Spectronaut. After importing the ion library through the search results, corresponding DIA files were loaded into Spectronaut and the aforementioned ion library was specified as reference identification database. The following custom settings were applied in Spectronaut: Calibration Mode: Forced. All results were filtered by a Q value of 0.01 (equals a FDR of 1% on peptide level). The output report was adjusted to contain all quantification related values. All other settings were set to default (by Spectronaut).

6.2.5 Statistical Analysis

Quantitative values were statistically analyzed with either SPSS v23 (IBM) or EXCEL with "Real Statistics Resource Pack" plugin. Further statistical analysis was performed with the proteomics data analysis tool: InfernoRDN (v.1.1.6044) ¹⁷⁴ from <http://omics.pnl.gov> (U.S. Department of Energy).

6.3 Results and Discussion

6.3.1 Sample Cohort

This sample cohort consists of a total of 36 samples which in turn can be divided into five distinct clinical distinctive groups: healthy controls, i.e. normal conditions and 4 different neurodegenerative diseases. Here, six samples represent the healthy state or normal physiological and psychological conditions. Furthermore, the neurodegenerative disease “*progressive supranuclear palsy*” (PSP) is represented by six samples. Morbus Parkinson or Parkinson's Disease (PD) as well as Morbus Alzheimer or Alzheimer's Disease (AD) are both represented by six samples each. Additionally, the progressive neurodegenerative disease “Corticobasal Degeneration” (CBD) is represented by a total of 12 samples. Due to inconsistency within the metadata in relation to the sample IDs, one sample that “apparently” belongs to the CBD group was however excluded from any data analysis. The focus of this study was the discovery of possible biomarker candidates for each of the neurodegenerative diseases. As some diseases feature overlapping molecular events as the cause of the diseases, this setup also allows for the discovery of stratifying biomarker candidates in cerebrospinal fluid (CSF).

6.3.2 Protein Quantification and Ion Library

For the ion library a total of 36 DDA files was analyzed via MaxQuant and its implemented Andromeda search algorithm. Hereby, feature matching was activated so that all possible precursor ions would be listed in the results file. After search result import, the ion library featured a total of 359 protein groups. This number correlates to 421 single proteins. Furthermore, 2329 peptides were saved in this ion library, which in turn were represented by 2791 precursors. Across the dataset an average of 344 ± 4 proteins was quantified per sample. However, in 2014, Guldbrandsen *et al.* ²²⁶ published a comprehensive analysis of CSF with different techniques and claimed that CSF in general features about 3000 different proteins. In comparison, this study hence only maps about 10% of the whole CSF proteome via mass spectrometry. Furthermore, it maintains the opinion that biomarker candidates might rather be found within low abundant proteins, which in turn are only observable after depletion of high abundant proteins ²²⁷⁻²³⁰. However, depletion of highly abundant proteins in CSF was waived for this study. Nonetheless, here, the protein “Serum Albumin” (P02768) was excluded from the analysis and hence removed from the list of quantified proteins before statistical analysis was applied. Especially in CSF, Albumin is considered the most abundant protein ²³¹ for this sample type. Nonetheless, corresponding Q-Q-plots for all samples (*cf.* **Figure 8.5.1**) showed that the assumption of an existent normal distribution can be rejected, as none of the plots featured a diagonal line as graphical outcome.

6.3.3 PSP-centered Biomarker Discovery

6.3.3.1 Discovery of Diagnostic Biomarkers for PSP

The cause of PSP is not yet fully understood. It is postulated that the establishment of the progressive supranuclear palsy disease involves the microtubule associated protein tau (P10636). In this context, the inherited genes which are passed on by the parents and are encoding for this protein appear to be affected.

Here, after quantifying proteins in CSF of healthy as well as PSP diagnosed patients a first investigation was a PC (principal component) analysis. Hereto, all proteins, which were previously quantified were used as input. Consequently, the initial PCA plot (cf. **Appendix - Figure 8.5.2**), including PC1 (43.3%) and PC2 (14.7%), accounted for 58% of the dataset variance. Also, the first trend of group-based separation was apparent as PSP samples spread along PC1 while healthy patients spread along PC2. However, an absolute clear separation was not achieved.

Consequently, proteins with significant protein abundance changes were identified by comparing the protein abundance of PSP samples to Controls (healthy patients). Here, a total of 11 proteins were identified, featuring a p-value (non-parametric; Mann-Whitney u test) less than 0.05 ($p \leq 0.05$). At this stage, a multiple testing correction was waived. Entering only these proteins into an according PCA analysis revealed a clear group-wise sample separation (cf. **Figure 6.1**). Furthermore, the corresponding PCA Bi-Plot demonstrates that the following two proteins are causing the disease specific spread of samples in this plot: Alpha-1-antichymotrypsin (P01011) and Complement component C9 (P02748).

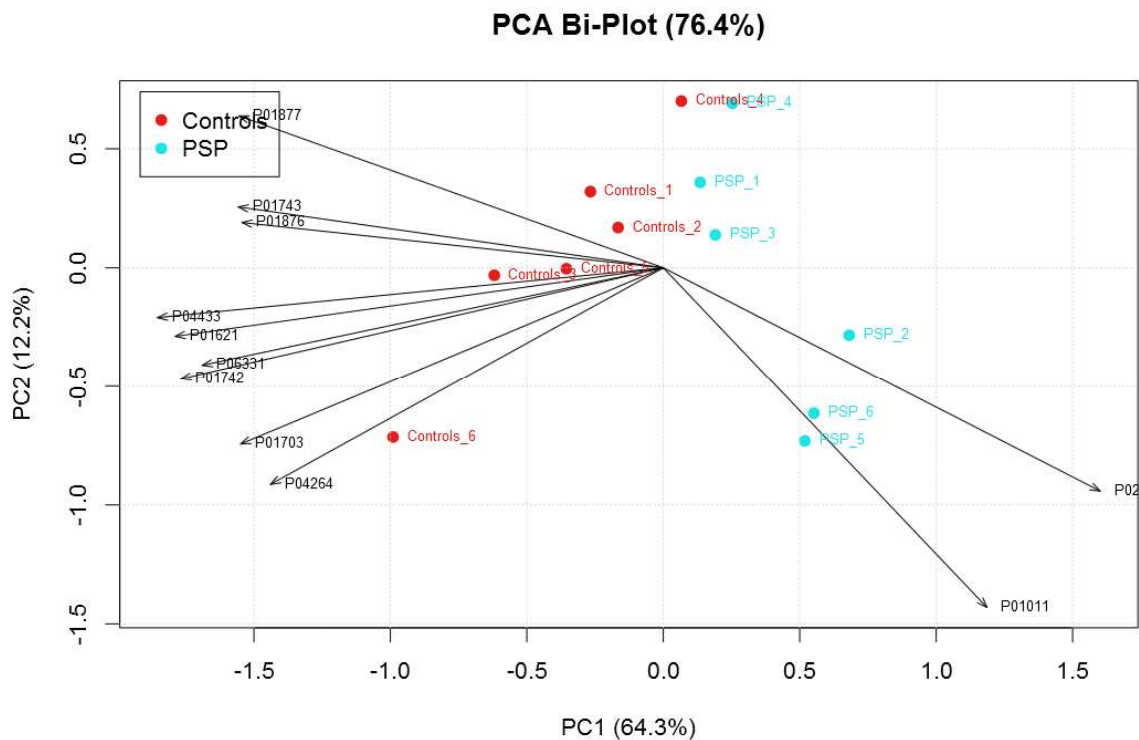


Figure 6.1: PCA Bi-Plot for significantly observed proteins (Controls vs. PSP). The first two principal components: PC1 (64.3 %) and PC2 (12.2 %) account for a total dataset variance of 76.4 %. Clear group-wise separation is apparent along PC1. The two features P01011 and P02748 particularly cause PSP separation.

In the context of neurodegeneration, both proteins were previously described as being involved in molecular process which lead to neurodegeneration or neuronal inflammation processes. Here, Alpha-1-antichymotrypsin is described as either being a key player of tau-phosphorylation in AD

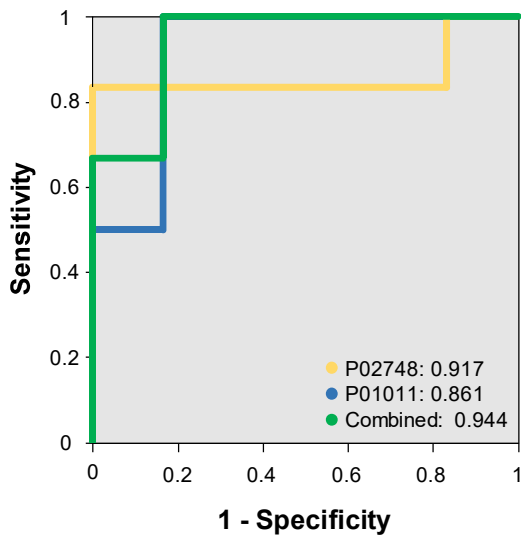


Figure 6.2: ROC curve for disease biomarker. Both proteins were tested for their state predictability. Complement component C9 (P02748) achieved an AUROC of 0.917 while Alpha-1-antichymotrypsin (P01011) only achieved 0.861. The linear combination of both achieved an AUROC of 0.944. Perfect biomarker conditions are represented by an AUROC of 1.

0.75. P02748 (Complement Component C9) featured 0.917, while P01011 (Alpha-1-antichymotrypsin) only scored 0.861. Nonetheless, the linear combination (binary regression) of both proteins and hence the addition of the predictability of these proteins featured the best AUROC of 0.944.

²³²⁻²³⁴ or was already found as a possible biomarker candidate for neurodegeneration ^{235, 236}. Furthermore, the Complement Component C9 protein is described as one of those proteins leading to complement activation and hence neuro-inflammation in AD ²³⁷⁻²³⁹. The described involvements however, are only in relation to Alzheimer's Disease and not PSP. Hence, these molecules might be of particular interest for the detection of general neurodegeneration in CSF.

Here, both proteins were further evaluated in terms of their diseased state predictability (rule-in). Therefore, the abundance information was entered into the calculation of a Receiver-Operator Characteristic curve (*cf.* **Figure 6.2**). For evaluation of their predictability the AUROC was deviated by ROC curve integration. Both discovered proteins featured an AUROC of >

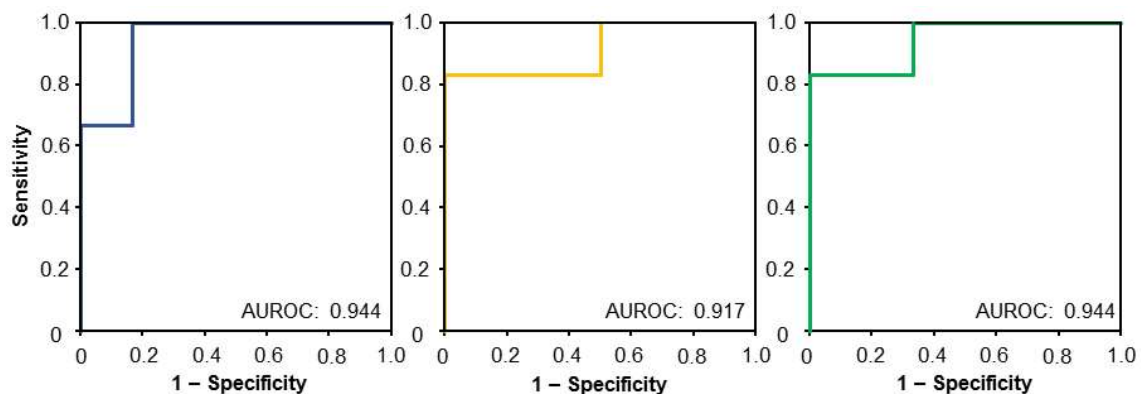


Figure 6.3: ROC curve for healthy biomarker candidates. Both proteins P01742 (blue) and P01743 (yellow) were tested for the diagnosing efficiency of a healthy state. In both cases the AUROC is greater than 0.9. Furthermore the linear combination of both (green) features an AUROC of 0.944. The linear combination did not achieve perfect biomarker panel conditions with an AUROC of 1.

In contrast to specifically diagnose the diseased state (rule-in), also the healthy state can be defined and hence diagnosed (rule-out). In this investigation, nine proteins (*cf.* **Figure 6.1**) show tendency to feature the ability to define the healthy state. Consequently, these proteins were analyzed regarding their solely diagnosing ability (ROC analysis). This analysis revealed that two proteins, namely P01742 (Immunoglobulin heavy variable 1-69) and P01743 (Immunoglobulin heavy variable 1-46), featured an AUROC of > 0.9 and hence can inherit diagnosing ability for the healthy state (*cf.* **Figure 6.3**). Additionally to the analysis of the proteins alone, also the linear combination of these was tested in regard of a ROC curve. Here, the AUROC numbers at 0.944. This however does also not represent perfect biomarker panel conditions with an AUROC of 1.

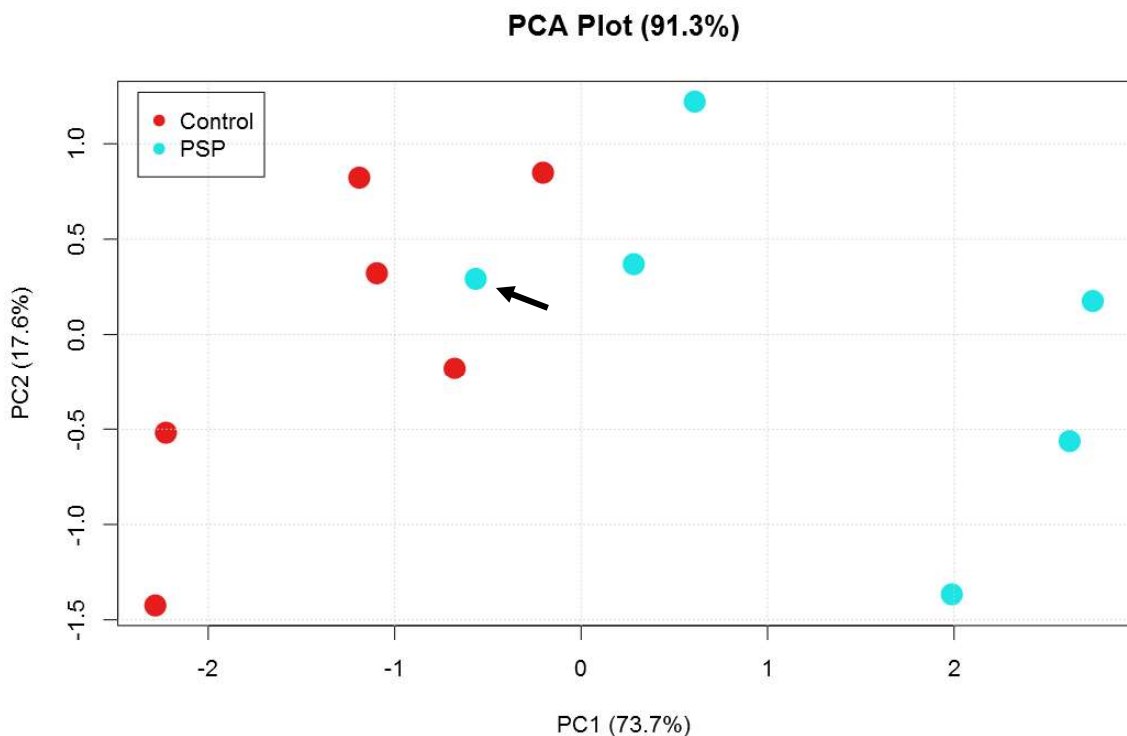


Figure 6.4: PCA plot of linear combined biomarker panel. PCA plots is based on the four identified proteins, which showed either diseases related or healthy state related segregation tendencies. The ROC analysis featured an AUROC of 1 while in the PCA plot one PSP sample (arrow) clusters very closely to healthy stated samples. Nonetheless, diagnosis can be easily deviated.

Ultimately, the identified four potential biomarker candidates (P01742, P01743, P02748 and P01011) were combined by linear regression. This linear regression allows for the combination of the diagnostic efficiency of all proteins at the same time. In this regard, the combined efficiency scored an AUROC of 1, which in turn represents perfect biomarker panel conditions. Additionally, in the PCA analysis while only these four proteins are considered, a clear separation was achieved (*cf.* **Figure 6.4**). Only one sample clustered closely with healthy stated samples.

Given these results, it can be concluded that the discovered proteins in this analysis are valuable biomarker candidates in the context of diagnosing neurodegeneration and PSP in particular. However, these biomarker candidates first need further validation, especially by methodologies other than mass spectrometry, such as Western Blot or ELISA, before future applications of these proteins as biomarker panel can be discussed.

6.3.3.2 Discovery of Stratification Biomarker for PSP

As mentioned before, there are four different neurodegenerative diseases represented in this dataset. Instead of simply investigating the dataset for diagnostic biomarkers, stratification-driven analyses can also be performed. Hence, instead of only comparing PSP diseases CSF samples with those samples originating from healthy individuals, here, PSP samples are also compared to those samples from the other three neurodegenerative diseases.

The calculation of the statistically significant abundance differences revealed a total of seven proteins, which demonstrated a p-value of < 0.05 . The corresponding PCA plot (cf. **Figure 6.5**), which in turn is only based on the previously discovered significant proteins demonstrates a group-wise separation. Here, all but one PSP sample separated from the other diseased samples.

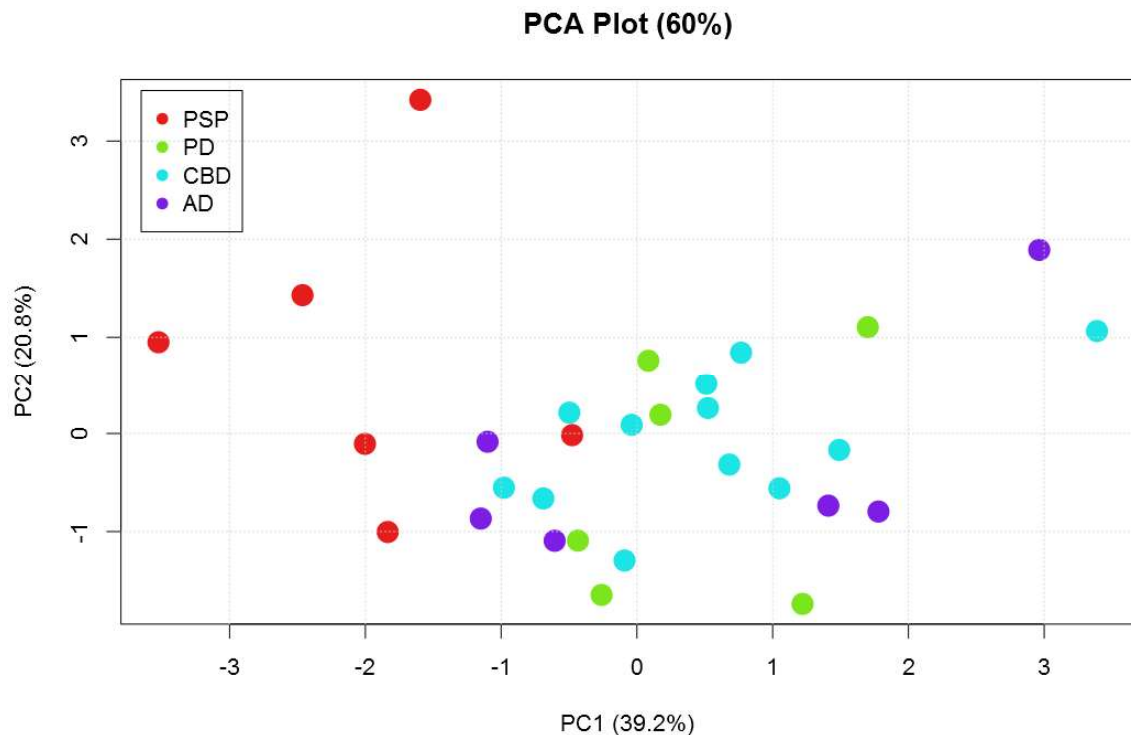


Figure 6.5: PCA plot for stratification. PCA plot for proteins with significant abundance changes while comparing PSP to other neurodegenerative diseases. The first two principal components account for 60 % of the dataset variance. All but one PSP sample separate from the other diseased CSF samples.

The corresponding PCA Bi-plot (cf. **Figure 8.5.3**) demonstrates, that the group-specific spreading is caused by a single protein: Q66K66 (Transmembrane protein 198). Furthermore, this protein has been described as being involved in the *Wnt* signaling pathway. Hereto, it is postulated that the transmembrane protein 198 causes LRP6 phosphorylation and also plays a role in the self-aggregation of LRP6. However, no link to neurodegenerative diseases could be found in the current literature. Hence, the transmembrane protein 198 (Q66K66) can be considered a possible stratification biomarker candidate, however further validation is needed.

6.3.4 PD-centered Biomarker Discovery

6.3.4.1 Discovery of Diagnostic Biomarkers

Parkinson's Disease (PD) is characterized as the neuronal degeneration of dopaminergic neurons in the substantia nigra ¹⁰⁴. Consequently, through the stalling of dopaminergic neurons the movement of affected patients is impaired and shows disease-specific symptoms such as shaking, rigidity, slowness of movement, and difficulty with walking ¹⁰⁵.

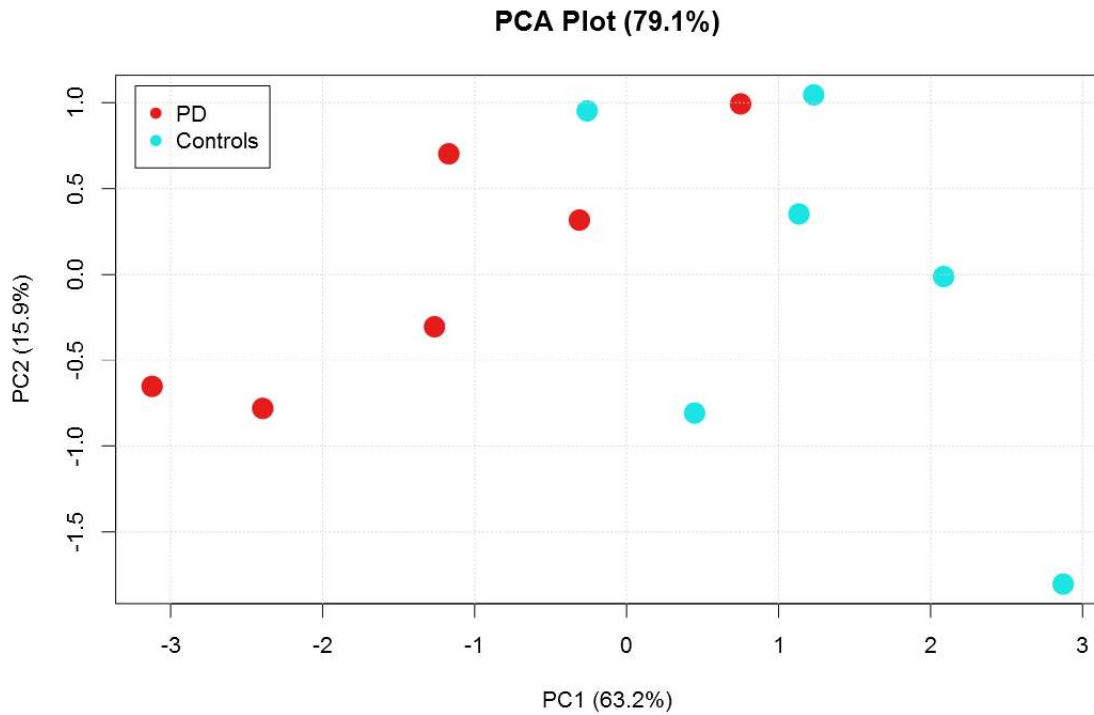


Figure 6.6: PCA plot for diagnosing PD in CSF. PCA plot, based on significantly different abundant proteins in CSF. First two principal components: PC1 (X – axis: 63.2 %) and PC2 (Y – axis: 15.9%) account for a total dataset variance of 79.1 %. One PD sample clusters closely to the Control samples (arrow). Segregation pattern can be deviated.

The direct comparison of diseased samples versus healthy patients revealed a total of five proteins that showed statistical significant abundance changes ($p < 0.05$). Here, multiple testing correction was waived. Based on the identified proteins a consecutive PCA analysis was performed. In this corresponding plot a group-wise separation is apparent. Nonetheless, one PD sample clusters very close to the cluster of control samples (*cf.* **Figure 6.6**).

Table 6.2: List of proteins observed with significant abundance changes.

Accession	Gen Tag	AUROC	p-Value
O00533	CHL1	0.917	0.0163091
P01023	A2M	0.889	0.0249746
O95025	SEMA3D	0.861	0.0373729
P08123	COL1A2	0.861	0.0373729
P13521	SCG2	0.861	0.0373729

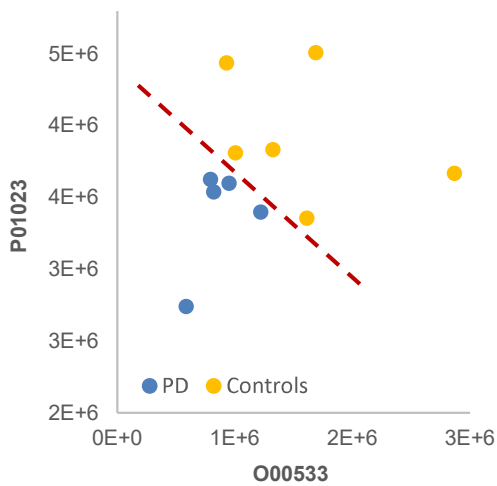


Figure 6.7: 2D scatter plot for PD and Controls. Plotted are observed protein abundances for O00533 (x-axis) and P01023 (y-axis). Simply the protein abundance of these two proteins allows for correct diagnosis (indicated by red dashed line).

In this case, perfect biomarker conditions are represented by a nominal AUROC of 1 and a clear group-wise separation in a 2D-scatterplot (cf. **Figure 6.7**). Nonetheless, similar conditions via linear regression of other features could not be achieved. Hence, it can be concluded, that the proteins O00533 and P01023 inherit biomarker panel potential. Nonetheless, further validation, as through Western Blot or an extended sample cohort is necessary before further conclusions can be stated.

As additional analysis, the discovered proteins with statistical significant protein abundance ($p < 0.05$) were entered into a ROC analysis. Here, all of the proteins featured an AUROC of greater than 0.85 (cf. **Table 6.2**). This in turn allows the assumption that all five proteins might be considered as possible biomarker candidates. Furthermore, two proteins, namely O00533 (Neural cell adhesion molecule L1-like protein) and P01023 (Alpha-2-macroglobulin) demonstrated the highest AUROC (AUROC > 0.88) amongst all tested. Based on the PCA Bi-plot (cf. **Figure 8.5.4**) these two proteins show no identical segregation potential. Hence, the feature effects were combined, here, through binary logistic regression and their combined effect tested again for predictability. In this

6.3.4.2 Discovery of Stratification Biomarker Candidates

The discovery of stratification biomarker candidates is again based on the statistical investigation for proteins that show significant abundance differences ($p < 0.05$) while comparing corresponding protein abundances on a group-wise basis. For stratification, protein abundances are directly compared to other neurodegenerative diseases implemented in this study.

The aforementioned comparison lead to the discovery of four proteins, which featured the required statistical significance. Nonetheless, multiple testing correction was waived in this context. Consequently, the observed proteins were entered into a PCA to observe the clustering ability in the context for stratification. In the corresponding plot, a certain clustering pattern could be deviated. However, perfect clustering conditions could not be achieved. Further, out of the mentioned PCA Bi-plot it was deviated that major parts of the PD-based clustering is based on two proteins: Q12888 (Tumor suppressor p53-binding protein 1) and P52701 (DNA mismatch repair protein Msh6). Based on these findings, the previously-discovered proteins were further investigated via ROC analysis.

In the ROC analysis, both proteins featured an AUROC of > 0.75. In particular, the Tumor suppressor p53-binding protein 1 (Q12888) scored a AUROC of 0.764 and DNA mismatch repair protein Msh6 (P52701) achieved 0.806 out of a total of 1. As apparent out of the previous PCA Bi-plot (cf. **Figure 6.8**), the effect vector for both proteins is pointing into the same direction.

Hence, the stratification ability was further investigated for the combined vector (Linear Binary Regression). Nonetheless, although both effect vectors were combined, considering both protein vectors at the same time only scored an AUROC of 0.833. These results indicate that these proteins feature a certain stratification tendency, but furthermore, don't feature the full potential to clearly differentiate between PD and other neurodegenerative diseases.

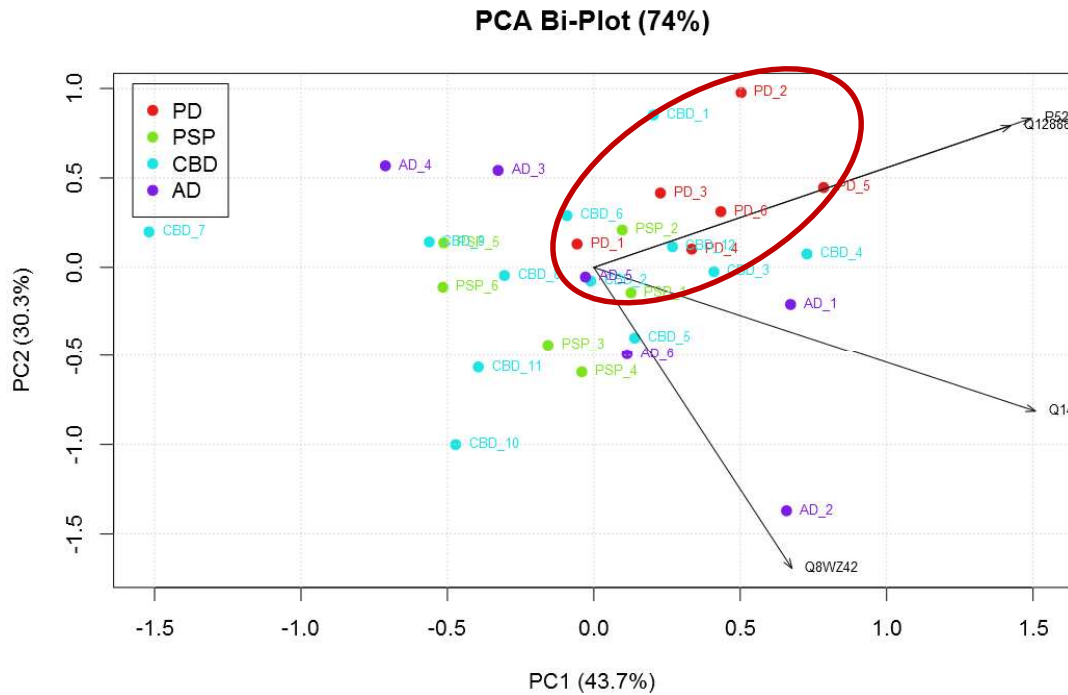


Figure 6.8: PCA Bi-Plot for biomarker candidates for PD stratification. The plot accounts for 74 % of the underlying dataset variance. Black arrows indicate PCA loadings. PD samples segregate along the Q12888 and P52701 vectors. No clear PD stratification can be deviated.

Based on the biological standpoint, both proteins are closely involved in the DNA repair mechanism. While Q12888 is especially involved in the binding of TP53 and the corresponding DNA damage signaling²⁴⁰, P52701 is described as part of the post-replicative DNA mismatch repair system (MMR)^{241, 242}. Most notably based on the fact that biomarkers are not necessarily directly associable with disease-specific pathway or biological processes²⁴³, here, both proteins were already described as being in the context of neurodegenerative diseases. In short, TP53BP1 (Q12888) was described as co-localizing with senataxin, especially in the context of ataxia. Yuce & West (2003)²⁴⁴ described the co-localization of these proteins and their signaling in DNA damage. Furthermore, Kulkarni & Wilson (2008)²⁴⁵ describe that the oxidative stress as pathological process in neurodegenerative disease leads to DNA damage and hence disease associated process can be classified for DNA repair.

Given these results, it can be concluded that two possible biomarker candidates have been discovered, as these proteins showed statistical significance as well as having already been described in literature as being involved in neurological diseases. Nonetheless, this approach is limited by sample numbers, which in turn might result in rejection as a candidate in larger or other sample cohorts. Nonetheless, further validation of the here identified proteins as biomarker candidates or biomarker panel is inevitably necessary.

6.3.5 AD-centered Biomarker Discovery

6.3.5.1 Discovery of Diagnostic Biomarker Candidates

Morbus Alzheimer or Alzheimer's disease (AD) describes the most common neurodegenerative disease with a prevalence of 60 to 70% of dementia cases. The specific target here, is the identification of biomarker candidates, which hold diagnostic value to particularly detect the early onsets of this disease.

Here, the protein abundances in CSF samples originating from patients suffering from Alzheimer's disease (AD) were statistically compared to those protein abundances originating from patients in a healthy (normal) state. This statistical comparison revealed a total of two proteins that showed significant ($p < 0.05$) changes. In particular, P01011 (Alpha-1-antichymotrypsin) and A6NNM8 (Tubulin polyglutamylase TTLL13P). In this context, Alpha-1-antichymotrypsin has already been identified with statistically significant protein abundance changes with the comparison of PSP samples with healthy state samples (*cf.* **Section 6.3.3.1**).

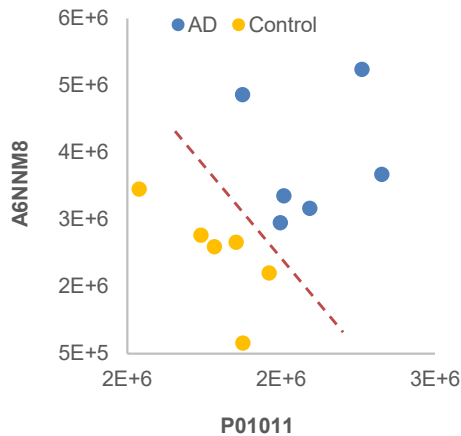


Figure 6.9: 2D scatter plot for AD diagnosis. The plot is based on the protein abundance observed for P01011 (x-axis) and A6NNM8 (y-axis). Separation tendency between AD and controls is indicated by a dashed red line

Given these results, it can be concluded that valuable biomarker candidates in CSF for the diagnosis of Alzheimer's Disease has been discovered. Nonetheless, further qualification and validation is inevitably required before further conclusions can be drawn. Also, given the small sample size (six versus six), the waived multiple testing correction, the chance of the outcome being biased and hence further, either unbiased or targeted validation is recommended.

The two identified proteins cannot be entered in a PCA analysis as too few features are available to calculate the corresponding principal components. Nonetheless, a group-wise separation can be deviated by displaying the proteins in a 2D scatter plot (*cf.* **Figure 6.9**). In this context, Alpha-1-antichymotrypsin is widely described as a possible marker in the context of Alzheimer's Disease ^{232, 246, 247}. Furthermore, it is mentioned in the pathophysiology of AD that patients suffering from AD feature decreased levels of α -tubulin. Additionally, it is described that α -tubulin is particularly post-translational modified. Here, the molecular function of Tubulin polyglutamylase (A6NNM8) is described as the transfer of "polyglutamyl" groups

onto tubulin. This process has been described in the context of AD itself, but also as part of the neuronal system in general ²⁴⁸⁻²⁵⁰.

Given these results, it can be concluded that valuable biomarker candidates in CSF for the diagnosis of Alzheimer's Disease has been discovered. Nonetheless, further qualification and validation is inevitably required before further conclusions can be drawn. Also, given the small sample size (six versus six), the waived multiple testing correction, the chance of the outcome being biased and hence further, either unbiased or targeted validation is recommended.

6.3.5.2 Discovery of Stratification Biomarker Candidates

The discovery of proteins for the diagnosing AD in CSF was previously described. Nonetheless, besides the strong need of diagnosing AD at the early onset of this disease, it is also important to be able to stratify between different neurological disorders. Hence, besides the comparison of CSF of AD patients with CSF samples of healthy patients, the acquired protein abundances were also compared to those of different neurological disorders (e.g. PSP, PD and CBD). This direct comparison, based on their diseased group affiliation, a total of seven proteins were identified that showed significant protein abundance changes ($p < 0.05$). Here, the multiple testing correction was also waived.

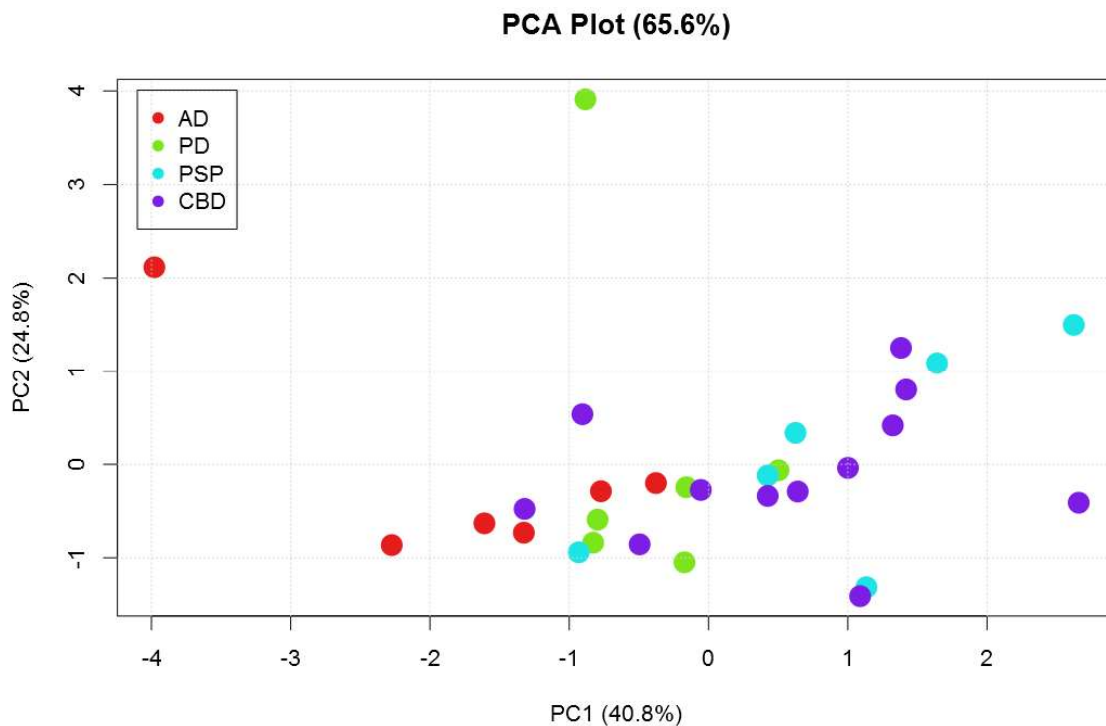


Figure 6.10: PCA plot for AD stratification. The PCA plot is based on the identified seven significantly different proteins. Here, the first two principal components account for 65.5 % of the dataset variance. Here, a slight segregation tendency can be deviated.

Consequently, all seven previously identified proteins were entered into a supervised PCA analysis. Nonetheless, no absolute clear group-based clustering or separation is apparent in the corresponding PCA plot (*cf.* **Figure 6.10**). However, a certain segregation pattern can be observed in the congruent PCA Bi-plot (*cf.* **Figure 8.5.5**). Here, the AD samples segregate themselves from the other ND samples along three particular proteins: Q96RW7 (Hemicentin-1), P02765 (Alpha-2-HS-glycoprotein) and P29622 (Kallistatin). These proteins are particularly deviated by the consideration of only the first two principal components (PC1 and PC2) which in turn account for 65.5 % of the significant protein dataset variation. The expansion of this consideration by the third component (PC3), out of the PCA Bi-plot considering PC1 and PC3, it can be deviated that Q9Y5Y7 (Lymphatic vessel endothelial hyaluronic acid receptor 1) also contributes to the stratification effect.

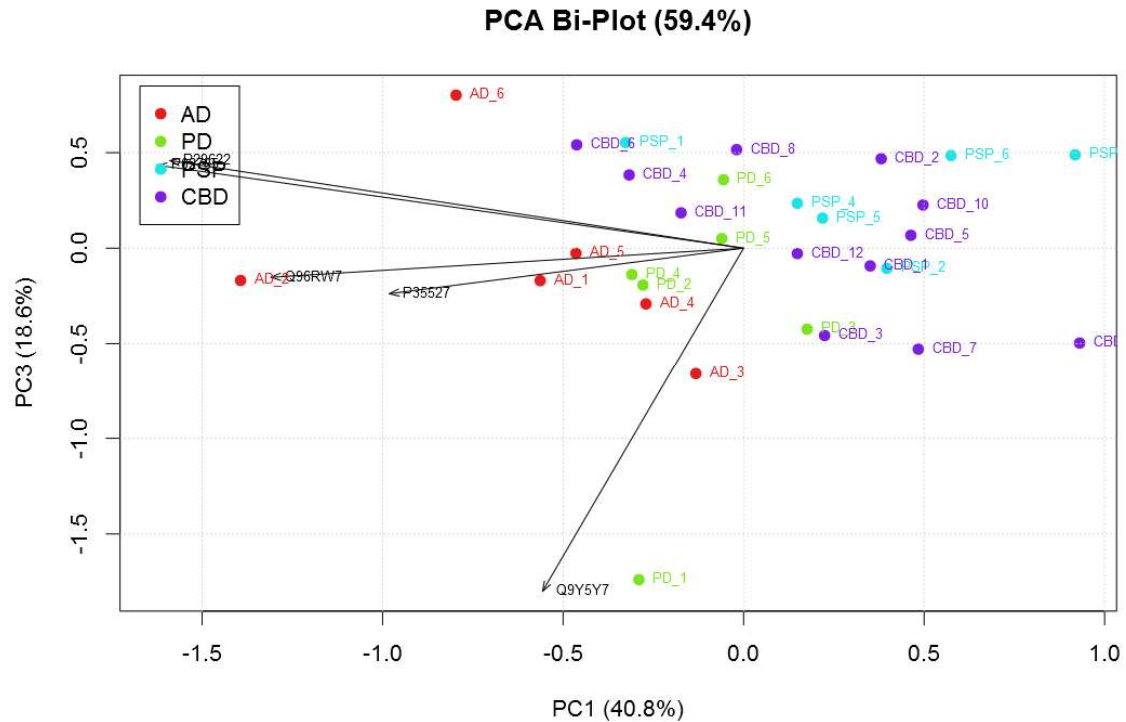


Figure 6.11: Extended PCA Bi-Plot for AD stratification. Instead of displaying PC1 and PC2, this PCA Bi-plot features the first (PC 1) and third (PC 3) principal component. The plot accounts for 59.4 % of the variance in the dataset of the previously identified seven proteins. Here, PC 3 accounts for 18.6 % variance.

A consequent literature review revealed that Alpha-2-Heremans-Schmid glycoprotein (AHSG) and Kallistatin/SERPINA4 had already been briefly described as a possible biomarker in the context of Alzheimer's Disease. In short, Geroldi *et al.* (2005)²⁵¹ investigates AHSG in the context of late onset Alzheimer's patients and the specific correlation to the genetic allele distribution. Their study found a notable correlation between allelic distribution and late onset of AD. Furthermore, Heit *et al.* (2013)²⁵² reports that serpins can generally influence protein aggregation. Additionally, it was mentioned in this context that SERPINI1 has been particularly correlated with dementia²⁵³, SERPINA5 being present in plaques in Multiple Sclerosis²⁵⁴ and SERPINA3 polymerization being an indicator for accelerated onset as well as severe progression of AD²⁵⁵. Furthermore, in a broader sense, the "Lymphatic vessel endothelial hyaluronic acid receptor 1" (LYVE-1) protein is used as a positive control for the presence of lymphatic tissue. In this regard and CSF is also considered as a lymphatic fluid and hence LYVE-1 is broadly used as control marker for CSF^{256, 257}.

As further analysis, all proteins were entered into ROC analysis. Here, all seven proteins scored an AUROC of greater than 0.7. In theory, based on this outcome, all seven proteins qualify as possible stratification biomarker candidates. Nonetheless, four proteins out of the initial seven proteins, namely Hemicentin-1 (Q96RW7), Joubertin (Q8N157), Kallistatin/SERPINA4 (P29622) and Alpha-2-HS-glycoprotein (P02765) demonstrated an AUROC of greater than 0.8. Here, the protein information of these proteins were combined via binary logistic regression and further tested by ROC analysis. The combined protein information achieved a total AUROC of 1 which equals perfect biomarker conditions.

Table 6.3: List of the identified seven stratification biomarker candidates.

Accession	Gene Tag	AUROC	p-Value
Q96RW7	HMCN1	0.854	0.008187
Q8N157	AHI1	0.847	0.009531
P29622	SERPINA4	0.840	0.011068
P02765	AHSG	0.833	0.012821
P02042	HBD	0.785	0.033524
P35527	KRT9	0.778	0.038088
Q9Y5Y7	LYVE1	0.771	0.043171

As previously mentioned, the following proteins were already described as possible biomarker candidates or marker in the context of CSF: P02765 (Alpha-2-HS-glycoprotein), P29622 (Kallistatin/SERPINA4) and Q9Y5Y7 (Lymphatic vessel endothelial hyaluronic acid receptor 1). For comparison these three proteins were also entered into the binary logistic regression to analyze their combined stratification potential. Nonetheless, only an AUROC of 0.910 was achieved. This value lies below perfect biomarker conditions but might be a valuable biomarker panel for the stratification of AD from other neurodegenerative diseases.

Given what has been previously described and discussed, these results demonstrate that valuable biomarker candidates or biomarker panels have been discovered in this sample cohort. Nonetheless, further extensive validation is necessary prior to be able to undertake further steps towards verification.

6.3.6 CBD-centered Biomarker Discovery

Corticobasal degeneration (CBD) is a rare progressive neurodegenerative disease which affects the cerebral cortex and basal ganglia. Like many neurodegenerative diseases, CBD also shares symptoms amongst major neurodegenerative diseases. Due to the various clinical presentations, CBD can only be finally diagnosed by neuropathological examination^{258, 259}. Hence, the discovery of biomarkers for either diagnosis or stratification remains a need that is yet unmet.

6.3.6.1 Discovery of Diagnostic Biomarker Candidates

Besides the aforementioned discovery studies, here the CBD group is represented with a total of 12 samples. This is then twice as much as for the other neurodegenerative diseases. The group-wise comparison for statistically significant protein abundance changes revealed a list of 118 proteins with $p < 0.05$. As previously mentioned, multiple testing correction was also waived here. Firstly, an unsupervised PCA analysis was performed to observe possible underlying segregation tendencies. In this corresponding PCA plot, a CBD-based group-wise segregation can already be observed. All but two samples do cluster together, based on their group association. Here, the first principal component (PC1) already accounts for 40.3 % of the whole dataset variation. It is along this component that the disease-related segregation takes place (cf. **Figure 8.5.6**). The addition of further components (PC2 and PC3) adds to the total amount of dataset variance considered in a certain plot, however, it does not further contribute to any segregation process.

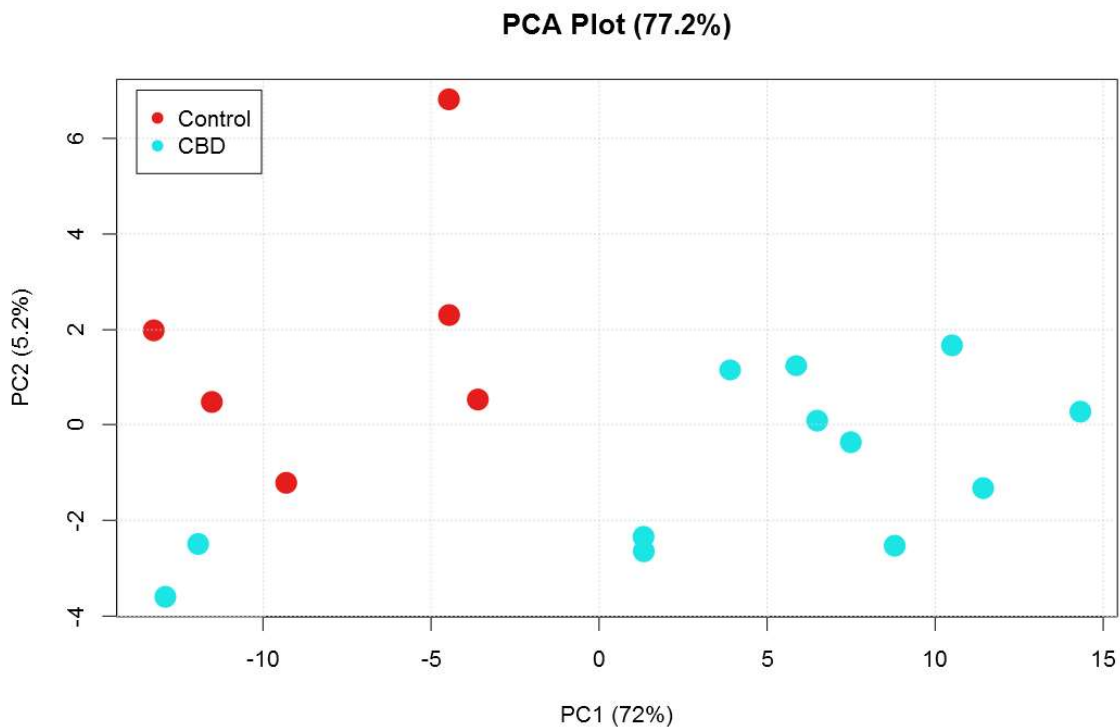


Figure 6.12: Supervised PCA plot for CBD biomarker discovery. This plot is based on the mentioned 118 significantly different proteins that were observed in the quantitative analysis. In this supervised PCA plot, the first component (PC 1) already accounts for 72% of the variation in this truncated dataset.

Furthermore, a supervised PCA analysis was performed to investigate the segregation efficiency of the different abundant proteins observed. Here, the aforementioned 118 proteins demonstrating a p-Value of less than 0.05 were entered into this analysis. The corresponding PCA plot demonstrates a clear group-wise segregation. Here, the two samples which did not cluster with the CBD group within the unsupervised analysis, are now perfectly separated from the control samples. Again, most of the group-wise segregation occurs across PC1, which in turn accounts for the majority of the variations observed (72 %). The previously falsely clustered CBD samples are further segregating along PC2, which respectively accounts for 5.2 % of the dataset variance and allow for the CBD samples to cluster beneath the control samples (*cf. Figure 6.12*).

Based on the corresponding PCA Bi-plot, only three distinct proteins cause the group-wise segregation effect. Here, only one feature (P00734, Prothrombin) demonstrates a strong effect strength along PC1 while the remainder (P01009, Alpha-1-antitrypsin and P01011, Alpha-1-antichymotrypsin) show major effect along PC2 (*cf. Figure 8.5.7*). As previously mentioned, certain Serpins (Serine proteases inhibitors) are considered biomarker candidates in the context of neurodegenerative diseases ^{252-254, 260}. In this context, Alpha-1-antitrypsin (SERPINA1) and Alpha-1-antichymotrypsin (SERPINA3) are members of the Serpin family. Furthermore, SERPINA1 has already been confirmed a valuable biomarker candidate in the context of PD ²⁶¹.

Table 6.4: List of the top three biomarker candidates for CBD.

Accession	Gene Tag	AUROC	p-Value
P00734	F2	0.889	0.0087302
P01011	SERPINA3	1.000	0.00074703
P01009	SERPINA1	0.806	0.0393518

Furthermore, the three previously identified proteins were evaluated regarding their diagnosing capability in the regard of Corticobasal degeneration. At first, the ROC curve for each protein was calculated alone. In general, each protein scored an AUROC of greater than 0.8. Alpha-1-antichymotrypsin (SERPINA3) scored an AUROC of 1, which in turn demonstrate perfect biomarker conditions. Besides SERPINA3, the linear combination out of two serpins that were found was also investigated regarding the predictability. The AUROC of latter linear combination also scored 1. Ultimately, the linear regression of all three proteins was investigated with a ROC analysis. This in turn also scored an AUROC of 1. Given these results, it can be concluded that SERPINA1, SERPINA3 and Prothrombin demonstrate valuable biomarker candidate proteins.

6.3.6.2 Discovery of Stratification Biomarker Candidates

In contrast to the Diagnostic Biomarker Discovery, the discovery of stratification biomarkers aims for the detection of proteins that help to differentiate between different neurodegenerative diseases based on the protein abundance changes. The discovery stage here, can be described as comparing of CSF protein abundances of variously diseased patients (AD, PD or PSP) versus those proteins abundances derived from those patients specifically suffering from CBD.

Based on the apparent non-normal distribution (i.e. non-parametric) of protein abundances across all examined patients the Mann-Whitney u test was used to compare the estimated protein abundances. With the application of this test to the existent dataset, a total of seven proteins that demonstrated a significant ($p < 0.05$) protein abundance change between the actual CBD and other neurodegenerative diseases were identified. Nonetheless, the CBD group here also features twice as many samples as those groups representing either AD, PD or PSP. A multiple testing correction of the calculated p-values was also waived here due to the small sample and protein list.

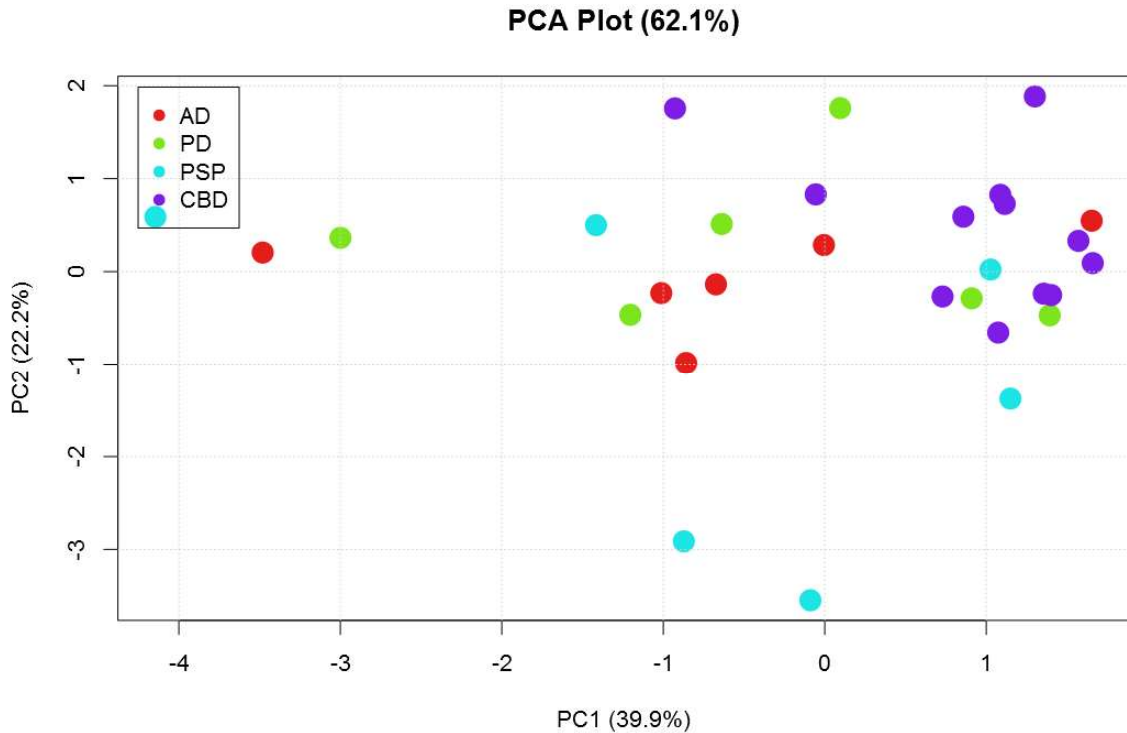


Figure 6.13: Supervised PCA plot for the discovery of stratification biomarker candidates. PCA plot is based on the seven identified significantly different abundant proteins. No absolute clear separation achieved. A certain cluster pattern can be deviated: CBD samples (purple) majorly group to the top right corner. PC1 (39.9 %) and PC2 (22.2 %) are plotted. A total dataset variance of 62.1 % is indicated in this plot.

Consequently, the identified significant proteins were entered into a PCA analysis (*cf.* **Figure 6.13**), which in turn reports the stratification capability of these proteins via group-wise clustering tendencies in the corresponding PCA plot. Here, no clear group-wise clustering was observed in the PCA plot. Nonetheless, a certain tendency was deviated, as CBD samples specifically accumulate in the top right corner of the plot (PC1 & PC2). Here, with the combination of PC1 and PC2, the plot accounts for 62.1% of the variation based on the seven aforementioned proteins. The alteration of the principal component combination, e.g. PC1 and PC3 or PC2 and PC3 (*cf.* **Appendix – Figure 8.5.8 ff.**), also didn't feature any absolute clear group-wise segregation. All plots only showed a certain tendency of segregation. Based on the PCA Bi-plots, which feature the certain feature loadings and hence allow for the detection of those feature causing the segregation, it could be deviated that CBD-based samples segregate across P06331 (Immunoglobulin heavy variable 4-34) and Q96RW7 (Hemicentin-1) in all three PC combinations, but also showed

segregation tendencies for Q66K66 (Transmembrane protein 198) in the PC1/PC3 plot. Since no clear clustering of CBD samples was achieved, all seven proteins were further evaluated.

After evaluation of the significantly different proteins in the PCA analysis, all of these proteins were consequently entered into the ROC analysis. Overall, all proteins demonstrated an AUROC of greater than 0.7, which however does immediately not qualify them as stratification biomarker candidates. Besides the investigation of the classification capability of proteins alone, certain protein combinations were also tested.

Table 6.5: List of the identified seven significant proteins.

Accession	Protein Name (Gene Tag)	AUROC
Q9Y6R7	IgGfC-binding protein (FCGBP)	0.796
Q12841	Follistatin-related protein 1 (FSTL1)	0.773
Q66K66	Transmembrane protein 198 (TMEM198)	0.759
P08779	Keratin, type I cytoskeletal 16 (KRT16)	0.738
P06331	Immunoglobulin heavy variable 4-34 (IGHV4-34)	0.736
Q96RW7	Hemicentin-1 (HMCN1)	0.731
P63261	Actin, cytoplasmic 2 (ACTG1)	0.718

In particular, the combination of P06331, Q96RW7 and Q66K66 was further investigated, as these proteins demonstrated the segregation tendencies in the PCA analysis. Here, this combination achieved an AUROC of 0.861, which clearly demonstrates better stratification conditions, however they are still not perfect conditions which would feature an AUROC of 1. Additionally, the combinations of the top two and top three scoring proteins in the single protein comparison were investigated. As result, the top two protein combination (Q9Y6R7 and Q12841) scored an AUROC of 0.808 while the top three protein combination (Q9Y6R7, Q12841 and Q66K66) scored an AUROC of 0.880. As already seen in the PCA analysis, non-perfect conditions were also only observed here within the ROC analysis (AUROC < 1).

A consequent thorough literature search for involvements of the seven discovered proteins however, did not provide any link or association with neurodegenerative processes or diseases. Furthermore, given the previously described results, it can be concluded that CBD might not be able to be further determined based on protein abundance changes in CSF of patients. Nonetheless, further qualification and/or validation has to be performed to deviate an eligible conclusion.

6.4 Conclusion and Perspective

Neurodegenerative disorders are mainly described by the loss of neuronal function or neuronal cell death. Symptoms mostly occur once the disease-specific processes are already running and neuronal degeneration has taken place to a certain extent. At the point of diagnosis of such a neurodegenerative disorder, the disease has most often progressed enormously, leaving the patients in an incurable state or with the only option being to try and slow down the degeneration process.

The aim for diagnostic biomarkers in the context of neurodegenerative diseases is the detection of a disease at its earliest onset. In this study, for the discovery of diagnostic biomarker candidates, CSF samples from healthy or non-diseased patients was compared to those CSF samples from patients who already obviously suffer from a specific neurodegenerative disorder. Hence, the observed biomarker candidates can already be correlated to disease-specific processes and thus allow for specific diagnosis. However, available samples which can clearly be assigned to a neurodegenerative disorder might not necessarily allow for the detection of “early onset” markers, which in turn are needed unconditionally. In this context, diagnostic markers should allow for the detection of the early onset, as well as for the specific classification of the disease. In this context, the limiting factor for a comprehensive discovery study depicts the availability as well as the extent of samples.

Besides diagnostic biomarkers which should help to detect the early onset of a certain disease, the stratification also imposes an important feature. In some cases, a final diagnosis can be given after pathophysiological examination of brain tissue. However, it is desirable to be able to easily differentiate between neurodegenerative disorders, especially since symptoms or symptom patterns can overlap with other diseases. Nonetheless, besides similar symptoms, the underlying processes can also easily overlap and hence leave the disease indistinguishable from others.

In this study, quantitative acquisition technologies were used for the discovery of diagnostic or stratification biomarker candidates. Indeed, valuable biomarker candidates were either newly discovered or confirmed through the data. All findings in this study are nevertheless based on a rather smaller cohort size ($N = 6$ or 12). Hence, the biomarker candidates found need further evaluation as well as verification in a much more comprehensive sample cohort than the present study. In this context, Skates *et al.* (2013)²⁶² postulates that given a sample in the type of a body fluid with a corresponding cohort size of 10 samples per group (i.e. 10 controls vs. 10 cases) and 20 markers planned for verification, each marker must be present in 80 % of the investigated samples with an SD difference of five-times to have an 80 % chance of actually reaching the final step of clinical verification. The extrapolation of this theory to this study implies that the biomarker candidates discovered here have a 1 - 8 % chance of entering clinical verification. Even though the biomarker candidates discovered here only have a low chance of entering clinical verification, they could confirm previously discovered biomarkers as well as provide insightful information about predominant molecular processes active with certain neurodegenerative diseases and thus help for stratification. Nonetheless, preliminary as well as pilot studies featuring a low number cohorts are an essential tool and are further required for further funding of large-scale studies.

6.5 Acknowledgements and Contributions

I want to thank Dr. Hanno Steen and Dr. Judith Steen for the opportunity to be able to work on this intriguing research project. Furthermore, I want to thank Dr. Jan Muntel for the initial input for the data-independent data acquisition methodology and support with questions arising during data analysis.

My personal contributions to this project account for: cohort management, sample preparation, carrying out the described experiments, analyzing the samples via data-independent mass spectrometric data acquisition, data analysis and figure generation.

7. References

1. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).
2. Pandey, A. & Mann, M. Proteomics to study genes and genomes. *Nature* **405**, 837-846 (2000).
3. Apweiler, R. et al. Approaching clinical proteomics: current state and future fields of application in cellular proteomics. *Cytometry A* **75**, 816-832 (2009).
4. Anderson, N.G. & Anderson, N.L. Twenty years of two-dimensional electrophoresis: past, present and future. *Electrophoresis* **17**, 443-453 (1996).
5. Celis, J.E. et al. Human 2-D PAGE databases for proteome analysis in health and disease: <http://biobase.dk/cgi-bin/celis>. *FEBS Lett* **398**, 129-134 (1996).
6. Wilkins, M.R. Proteome research : new frontiers in functional genomics. (Springer, Berlin ; New York; 1997).
7. Wilkins, M.R. et al. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **14**, 61-65 (1996).
8. Larance, M. & Lamond, A.I. Multidimensional proteomics for cell biology. *Nat Rev Mol Cell Biol* **16**, 269-280 (2015).
9. Ossipova, E., Fenyo, D. & Eriksson, J. Optimizing search conditions for the mass fingerprint-based identification of proteins. *Proteomics* **6**, 2079-2085 (2006).
10. Hillenkamp, F. & Karas, M. Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Methods Enzymol* **193**, 280-295 (1990).
11. Küster, B., Edn. Lecture Notes. (ed. L.f. Bioanalytik) (Technische Universität München, Freising; 2010).
12. Rehm, H. & Letzel, T. Der Experimentator: Proteinbiochemie/Proteomics, Edn. 6th. (Springer Spektrum, Heidelberg; 2010).
13. Kelleher, N.L. Top-down proteomics. *Anal Chem* **76**, 197A-203A (2004).
14. Wu, C. et al. A protease for 'middle-down' proteomics. *Nat Methods* **9**, 822-824 (2012).
15. Chait, B.T. Chemistry. Mass spectrometry: bottom-up or top-down? *Science* **314**, 65-66 (2006).
16. Moradian, A., Kalli, A., Sweredoski, M.J. & Hess, S. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics* **14**, 489-497 (2014).
17. Swaney, D.L., Wenger, C.D. & Coon, J.J. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* **9**, 1323-1329 (2010).
18. Bogdanov, B. & Smith, R.D. Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom Rev* **24**, 168-200 (2005).
19. Han, X., Aslanian, A. & Yates, J.R., 3rd Mass spectrometry for proteomics. *Curr Opin Chem Biol* **12**, 483-490 (2008).

20. Horváth, C. High-performance liquid chromatography. (Academic Press Inc., New York; 1983).
21. Wilm, M. et al. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466-469 (1996).
22. van Deemter, J.J., Zuiderweg, F.J. & Klinkenberg, A. Frontiers of Chemical Engineering Science Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chemical Engineering Science* **50**, 3869-3882 (1995).
23. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).
24. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299-2301 (1988).
25. Barber, M., Bordoli, R.S., Sedgwick, R.D. & Tyler, A.N. Fast atom bombardment mass spectrometry of the angiotensin peptides. *Biomed Mass Spectrom* **9**, 208-214 (1982).
26. Morris, H.R. et al. Fast atom bombardment: a new mass spectrometric method for peptide sequence analysis. *Biochem Biophys Res Commun* **101**, 623-631 (1981).
27. Macfarlane, R.D. & Torgerson, D.F. Californium-252 plasma desorption mass spectroscopy. *Science* **191**, 920-925 (1976).
28. Lottspeich, F. & Engels, J.W. Bioanalytik. (Spektrum Akademischer Verlag, Heidelberg; 2012).
29. Banerjee, S. & Mazumdar, S. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem* **2012**, 282574 (2012).
30. Clark, A.E., Kaleta, E.J., Arora, A. & Wolk, D.M. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clin Microbiol Rev* **26**, 547-603 (2013).
31. Jurinke, C., Oeth, P. & van den Boom, D. MALDI-TOF mass spectrometry: a versatile tool for high-performance DNA analysis. *Mol Biotechnol* **26**, 147-164 (2004).
32. Aebersold, R. & Goodlett, D.R. Mass spectrometry in proteomics. *Chem Rev* **101**, 269-295 (2001).
33. Mann, M., Hendrickson, R.C. & Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* **70**, 437-473 (2001).
34. Burlingame, A.L., Boyd, R.K. & Gaskell, S.J. Mass spectrometry. *Anal Chem* **70**, 647R-716R (1998).
35. Cotter, R.J. Time-of-flight mass spectrometry: an increasing role in the life sciences. *Biomed Environ Mass Spectrom* **18**, 513-532 (1989).
36. Uphoff, A. & Grotemeyer, J. Tutorial: the secrets of time-of flight mass spectrometry revealed. *Eur J Mass Spectrom (Chichester, Eng)* **9**, 151-164 (2003).
37. Makarov, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* **72**, 1156-1162 (2000).

38. Makarov, A. et al. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* **78**, 2113-2120 (2006).
39. Hu, Q. et al. The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **40**, 430-443 (2005).
40. Makarov, A., Denisov, E., Lange, O. & Horning, S. Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J Am Soc Mass Spectrom* **17**, 977-982 (2006).
41. Perry, R.H., Cooks, R.G. & Noll, R.J. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom Rev* **27**, 661-699 (2008).
42. McLafferty, F.W. & Tureček, F.e. Interpretation of mass spectra, Edn. 4th. (University Science Books, Mill Valley, Calif.; 1993).
43. Hunt, D.F., Yates, J.R., 3rd, Shabanowitz, J., Winston, S. & Hauer, C.R. Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A* **83**, 6233-6237 (1986).
44. Olsen, J.V. et al. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **4**, 709-712 (2007).
45. Kim, M.S. & Pandey, A. Electron transfer dissociation mass spectrometry in proteomics. *Proteomics* **12**, 530-542 (2012).
46. Syka, J.E., Coon, J.J., Schroeder, M.J., Shabanowitz, J. & Hunt, D.F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **101**, 9528-9533 (2004).
47. Sleno, L. & Volmer, D.A. Ion activation methods for tandem mass spectrometry. *J Mass Spectrom* **39**, 1091-1112 (2004).
48. Chi, A. et al. Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci U S A* **104**, 2193-2198 (2007).
49. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **11**, 601 (1984).
50. Qi, Y. & Volmer, D.A. Electron-based fragmentation methods in mass spectrometry: An overview. *Mass Spectrom Rev* (2015).
51. Trauger, S.A., Webb, W. & Siuzdak, G. Peptide and protein analysis with mass spectrometry. *Spectroscopy* **16** (2002).
52. Gupta, N. & Pevzner, P.A. False discovery rates of protein identifications: a strike against the two-peptide rule. *J Proteome Res* **8**, 4173-4181 (2009).
53. Aggarwal, S. & Yadav, A.K. False Discovery Rate Estimation in Proteomics. *Methods Mol Biol* **1362**, 119-128 (2016).
54. Reiter, L. et al. Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol Cell Proteomics* **8**, 2405-2417 (2009).
55. Savitski, M.M., Wilhelm, M., Hahne, H., Kuster, B. & Bantscheff, M. A Scalable Approach for Protein False Discovery Rate Estimation in Large Proteomic Data Sets. *Mol Cell Proteomics* **14**, 2394-2404 (2015).

56. Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J. & Gygi, S.P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* **2**, 43-50 (2003).
57. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **389**, 1017-1031 (2007).
58. Nikolov, M., Schmidt, C. & Urlaub, H. Quantitative mass spectrometry-based proteomics: an overview. *Methods Mol Biol* **893**, 85-100 (2012).
59. Ong, S.E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* **1**, 252-262 (2005).
60. Castillo, M.J., Reynolds, K.J., Gomes, A., Fenselau, C. & Yao, X. Quantitative protein analysis using enzymatic [(1)(8)O]water labeling. *Curr Protoc Protein Sci* **76**, 23 24 21-29 (2014).
61. Heller, M., Mattou, H., Menzel, C. & Yao, X. Trypsin catalyzed 16O-to-18O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers. *J Am Soc Mass Spectrom* **14**, 704-718 (2003).
62. Ye, X., Luke, B., Andresson, T. & Blonder, J. 18O stable isotope labeling in MS-based proteomics. *Brief Funct Genomic Proteomic* **8**, 136-144 (2009).
63. Gygi, S.P. et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-999 (1999).
64. Del Campo, M. et al. Facilitating the Validation of Novel Protein Biomarkers for Dementia: An Optimal Workflow for the Development of Sandwich Immunoassays. *Front Neurol* **6**, 202 (2015).
65. Butler, C. & Zeman, A.Z. Neurological syndromes which can be mistaken for psychiatric conditions. *J Neurol Neurosurg Psychiatry* **76 Suppl 1**, i31-38 (2005).
66. Purves, D. Neuroscience. (Sinauer Associates, Sunderland, Mass.; 2012).
67. Azevedo, F.A. et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol* **513**, 532-541 (2009).
68. Bahney, J. & von Bartheld, C.S. Validation of the isotropic fractionator: comparison with unbiased stereology and DNA extraction for quantification of glial cells. *J Neurosci Methods* **222**, 165-174 (2014).
69. West, M.J., Slomianka, L. & Gundersen, H.J. Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator. *Anat Rec* **231**, 482-497 (1991).
70. Zilles, K., Tillmann B. N. Anatomie. (Springer-Verlag, Berlin; 2010).
71. Zündorf, I. Anatomie und Physiologie. Von G.J. Tortora, B.H. Derrickson. *Pharmazie in unserer Zeit* **36**, 70-70 (2007).
72. Mackenzie, I.R. et al. Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: an update. *Acta Neuropathol* **119**, 1-4 (2010).

73. Finset, A. & Andersson, S. Coping strategies in patients with acquired brain injury: relationships between coping, apathy, depression and lesion location. *Brain Inj* **14**, 887-905 (2000).
74. Sohlberg, M.M., McLaughlin, K.A., Pavese, A., Heidrich, A. & Posner, M.I. Evaluation of attention process training and brain injury education in persons with acquired brain injury. *J Clin Exp Neuropsychol* **22**, 656-676 (2000).
75. Daneshvar, D.H., Nowinski, C.J., McKee, A.C. & Cantu, R.C. The epidemiology of sport-related concussion. *Clin Sports Med* **30**, 1-17, vii (2011).
76. Arciniegas, D.B., Anderson, C.A., Topkoff, J. & McAllister, T.W. Mild traumatic brain injury: a neuropsychiatric approach to diagnosis, evaluation, and treatment. *Neuropsychiatr Dis Treat* **1**, 311-327 (2005).
77. Flanagan, S.R., Cantor, J.B. & Ashman, T.A. Traumatic brain injury: future assessment tools and treatment prospects. *Neuropsychiatr Dis Treat* **4**, 877-892 (2008).
78. Clinic, M. in Diseases and Conditions (2014).
79. Maas, A.I., Stocchetti, N. & Bullock, R. Moderate and severe traumatic brain injury in adults. *Lancet Neurol* **7**, 728-741 (2008).
80. Zink, B.J. Traumatic brain injury outcome: concepts for emergency care. *Annals of emergency medicine* **37**, 318-332 (2001).
81. Mills, J.D., Bailes, J.E., Sedney, C.L., Hutchins, H. & Sears, B. Omega-3 fatty acid supplementation and reduction of traumatic axonal injury in a rodent head injury model. *J Neurosurg* **114**, 77-84 (2011).
82. Zonfrillo, M.R. et al. Pediatric providers' self-reported knowledge, practices, and attitudes about concussion. *Pediatrics* **130**, 1120-1125 (2012).
83. Kwasnica, C., Brown, A.W., Elovic, E.P., Kothari, S. & Flanagan, S.R. Congenital and acquired brain injury. 3. Spectrum of the acquired brain injury population. *Arch Phys Med Rehabil* **89**, S15-20 (2008).
84. Boutis, K., Weerdenburg, K., Koo, E., Schneeweiss, S. & Zemek, R. The diagnosis of concussion in a pediatric emergency department. *The Journal of pediatrics* **166**, 1214-1220 e1211 (2015).
85. Draper, K. & Ponsford, J. Long-term outcome following traumatic brain injury: a comparison of subjective reports by those injured and their relatives. *Neuropsychol Rehabil* **19**, 645-661 (2009).
86. Parikh, S., Koch, M. & Narayan, R.K. Traumatic brain injury. *Int Anesthesiol Clin* **45**, 119-135 (2007).
87. Bredesen, D.E., Rao, R.V. & Mehlen, P. Cell death in the nervous system. *Nature* **443**, 796-802 (2006).
88. Rubinsztein, D.C. The roles of intracellular protein-degradation pathways in neurodegeneration. *Nature* **443**, 780-786 (2006).
89. Thompson, L.M. Neurodegeneration: a question of balance. *Nature* **452**, 707-708 (2008).
90. Walker, F.O. Huntington's disease. *Lancet* **369**, 218-228 (2007).

91. Davie, C.A. A review of Parkinson's disease. *Br Med Bull* **86**, 109-127 (2008).
92. Lesage, S. & Brice, A. Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum Mol Genet* **18**, R48-59 (2009).
93. Chun, W. & Johnson, G.V. The role of tau phosphorylation and cleavage in neuronal cell death. *Front Biosci* **12**, 733-756 (2007).
94. Hashimoto, M., Rockenstein, E., Crews, L. & Masliah, E. Role of protein aggregation in mitochondrial dysfunction and neurodegeneration in Alzheimer's and Parkinson's diseases. *Neuromolecular Med* **4**, 21-36 (2003).
95. Bastian, F.O. et al. Spiroplasma spp. from transmissible spongiform encephalopathy brains or ticks induce spongiform encephalopathy in ruminants. *J Med Microbiol* **56**, 1235-1242 (2007).
96. Burns, A. & Iliffe, S. Alzheimer's disease. *BMJ* **338**, b158 (2009).
97. Huang, Y. & Mucke, L. Alzheimer mechanisms and therapeutic strategies. *Cell* **148**, 1204-1222 (2012).
98. Hooper, N.M. Roles of proteolysis and lipid rafts in the processing of the amyloid precursor protein and prion protein. *Biochem Soc Trans* **33**, 335-338 (2005).
99. Ohnishi, S. & Takano, K. Amyloid fibrils from the viewpoint of protein folding. *Cell Mol Life Sci* **61**, 511-524 (2004).
100. Priller, C. et al. Synapse formation and function is modulated by the amyloid precursor protein. *J Neurosci* **26**, 7212-7221 (2006).
101. Turner, P.R., O'Connor, K., Tate, W.P. & Abraham, W.C. Roles of amyloid precursor protein and its fragments in regulating neural activity, plasticity and memory. *Prog Neurobiol* **70**, 1-32 (2003).
102. Hernandez, F. & Avila, J. Tauopathies. *Cell Mol Life Sci* **64**, 2219-2233 (2007).
103. Nerius, M., Fink, A. & Doblhammer, G. Parkinson's disease in Germany: prevalence and incidence based on health claims data. *Acta Neurol Scand* (2016).
104. Kalia, L.V. & Lang, A.E. Parkinson's disease. *Lancet* **386**, 896-912 (2015).
105. Sveinbjornsdottir, S. The clinical symptoms of Parkinson's disease. *J Neurochem* **139 Suppl 1**, 318-324 (2016).
106. Hirsch, E.C. Iron transport in Parkinson's disease. *Parkinsonism Relat Disord* **15 Suppl 3**, S209-211 (2009).
107. Obeso, J.A. et al. Missing pieces in the Parkinson's disease puzzle. *Nat Med* **16**, 653-661 (2010).
108. Schulz-Schaeffer, W.J. The synaptic pathology of alpha-synuclein aggregation in dementia with Lewy bodies, Parkinson's disease and Parkinson's disease dementia. *Acta Neuropathol* **120**, 131-143 (2010).
109. Manza, L.L., Stamer, S.L., Ham, A.J., Codreanu, S.G. & Liebler, D.C. Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* **5**, 1742-1745 (2005).
110. Wisniewski, J.R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).

111. Switzar, L., van Angeren, J., Pinkse, M., Kool, J. & Niessen, W.M. A high-throughput sample preparation method for cellular proteomics using 96-well filter plates. *Proteomics* **13**, 2980-2983 (2013).
112. Yu, Y. et al. Urine sample preparation in 96-well filter plates for quantitative clinical proteomics. *Anal Chem* **86**, 5470-5477 (2014).
113. Naldrett, M.J., Zeidler, R., Wilson, K.E. & Kocourek, A. Concentration and desalting of peptide and protein samples with a newly developed C18 membrane in a microspin column format. *Journal of biomolecular techniques : JBT* **16**, 423-428 (2005).
114. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-254 (1976).
115. McKeon, T.A. & Lyman, M.L. Calcium ion improves electrophoretic transfer of calmodulin and other small proteins. *Anal Biochem* **193**, 125-130 (1991).
116. Schwanhauser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).
117. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).
118. Vizcaino, J.A. et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* **32**, 223-226 (2014).
119. Baker, P.R. & Chalkley, R.J. MS-viewer: a web-based spectral viewer for proteomics results. *Mol Cell Proteomics* **13**, 1392-1396 (2014).
120. Ivanov, A.R., Lazarev, A.V. Sample preparation in biological mass spectrometry, Edn. 1st. (Springer, New York; 2011).
121. Sloane, A.J. et al. High throughput peptide mass fingerprinting and protein macroarray analysis using chemical printing strategies. *Molecular & cellular proteomics : MCP* **1**, 490-499 (2002).
122. Methogo, R.M., Dufresne-Martin, G., Leclerc, P., Leduc, R. & Klarskov, K. Mass spectrometric peptide fingerprinting of proteins after Western blotting on polyvinylidene fluoride and enhanced chemiluminescence detection. *J Proteome Res* **4**, 2216-2224 (2005).
123. Fernandez, J. & Mische, S.M. Enzymatic digestion of proteins on PVDF membranes. *Curr Protoc Protein Sci* **Chapter 11**, Unit 11 12 (2001).
124. Eckerskorn, C. & Lottspeich, F. Structural characterization of blotting membranes and the influence of membrane parameters for electroblotting and subsequent amino acid sequence analysis of proteins. *Electrophoresis* **14**, 831-838 (1993).
125. Dickhut, C., Feldmann, I., Lambert, J. & Zahedi, R.P. Impact of digestion conditions on phosphoproteomics. *J Proteome Res* **13**, 2761-2770 (2014).
126. Lee, Y.H. & Song, G.G. Plasminogen activator inhibitor-1 4G/5G and the MTHFR 677C/T polymorphisms and susceptibility to polycystic ovary syndrome: a meta-analysis. *European journal of obstetrics, gynecology, and reproductive biology* **175**, 8-14 (2014).

127. Koiou, E. et al. Plasma plasminogen activator inhibitor-1 levels in the different phenotypes of the polycystic ovary syndrome. *Endocrine journal* **59**, 21-29 (2012).
128. Miyazaki, K. et al. Clinical usefulness of serum squamous cell carcinoma antigen for early detection of squamous cell carcinoma arising in mature cystic teratoma of the ovary. *Obstetrics and gynecology* **78**, 562-566 (1991).
129. Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **11**, 319-324 (2014).
130. Prevention, T.C.f.D.C.a. (National Center for Injury Prevention and Control; Division of Unintentional Injury Prevention. Atlanta, GA, USA, 2014).
131. Marin, J.R., Weaver, M.D., Yealy, D.M. & Mannix, R.C. Trends in visits for traumatic brain injury to emergency departments in the United States. *Jama* **311**, 1917-1919 (2014).
132. Yeates, K.O. in Pediatric neurophysiology: research, theory, and practice, Edn. 2nd. (ed. M.D.R. K. O. Yeates, H. G. Taylor, & B. F. Pennington) 112-146 (Guilford, New York; 2010).
133. Bodin, D., Yeates, K. O., Klamar K. in Pediatric and Adolescent Concussion: Diagnosis, Management and Outcomes. (ed. J.N.A.a.K.D. Walter) (Springer, New York; 2012).
134. Halstead, M.E., Walter, K.D., Council on Sports, M. & Fitness American Academy of Pediatrics. Clinical report--sport-related concussion in children and adolescents. *Pediatrics* **126**, 597-615 (2010).
135. McCrea, H.J., Perrine, K., Niogi, S. & Hartl, R. Concussion in sports. *Sports health* **5**, 160-164 (2013).
136. Mannix, R., Eisenberg, M., Berry, M., Meehan, W.P., 3rd & Hayes, R.L. Serum biomarkers predict acute symptom burden in children after concussion: a preliminary study. *J Neurotrauma* **31**, 1072-1075 (2014).
137. Zetterberg, H. et al. Neurochemical aftermath of amateur boxing. *Archives of neurology* **63**, 1277-1280 (2006).
138. Zetterberg, H. & Blennow, K. Fluid markers of traumatic brain injury. *Molecular and cellular neurosciences* **66**, 99-102 (2015).
139. Unden, J., Ingebrigtsen, T., Romner, B. & Scandinavian Neurotrauma, C. Scandinavian guidelines for initial management of minimal, mild and moderate head injuries in adults: an evidence and consensus-based update. *BMC medicine* **11**, 50 (2013).
140. Neselius, S. et al. CSF-biomarkers in Olympic boxing: diagnosis and effects of repetitive head trauma. *PloS one* **7**, e33606 (2012).
141. McMahon, P.J. et al. Measurement of the glial fibrillary acidic protein and its breakdown products GFAP-BDP biomarker for the detection of traumatic brain injury compared to computed tomography and magnetic resonance imaging. *J Neurotrauma* **32**, 527-533 (2015).
142. Shi, M. et al. Salivary tau species are potential biomarkers of Alzheimer's disease. *Journal of Alzheimer's disease : JAD* **27**, 299-305 (2011).

143. Adachi, J., Kumar, C., Zhang, Y., Olsen, J.V. & Mann, M. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome biology* **7**, R80 (2006).
144. Brenner, B.M. & Rector, F.C. Brenner & Rector's the kidney, Edn. 8th. (Saunders Elsevier, Philadelphia; 2008).
145. Brunzel, N.A. Fundamentals of urine & body fluid analysis, Edn. 3rd. (Elsevier/Saunders, St. Louis, Mo.; 2013).
146. Muntel, J. et al. Advancing Urinary Protein Biomarker Discovery by Data-Independent Acquisition on a Quadrupole-Orbitrap Mass Spectrometer. *J Proteome Res* **14**, 4752-4762 (2015).
147. 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* **14**, 1400-1410 (2015).
148. Eisenberg, M.A., Andrea, J., Meehan, W. & Mannix, R. Time interval between concussions and symptom duration. *Pediatrics* **132**, 8-17 (2013).
149. Rifai, N., Gillette, M.A. & Carr, S.A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* **24**, 971-983 (2006).
150. Fay, M.P. & Proschan, M.A. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Stat Surv* **4**, 1-39 (2010).
151. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
152. Lee, K., Lim, S., Bartell, S. & Hong, Y.C. Interpersonal and temporal variability of urinary cotinine in elderly subjects. *Int J Hyg Environ Health* **215**, 46-50 (2011).
153. Ghosh, D. & Poisson, L.M. "Omics" data and levels of evidence for biomarker discovery. *Genomics* **93**, 13-16 (2009).
154. Pepe, M.S., Li, C.I. & Feng, Z. Improving the quality of biomarker discovery research: the right samples and enough of them. *Cancer Epidemiol Biomarkers Prev* **24**, 944-950 (2015).
155. Segata, N. et al. Metagenomic biomarker discovery and explanation. *Genome biology* **12**, R60 (2011).
156. Shi, J., Longo, F.M. & Massa, S.M. A small molecule p75(NTR) ligand protects neurogenesis after traumatic brain injury. *Stem Cells* **31**, 2561-2574 (2013).
157. Sebastiani, A. et al. Proneurotrophin Binding to P75 Neurotrophin Receptor (P75ntr) Is Essential for Brain Lesion Formation and Functional Impairment after Experimental Traumatic Brain Injury. *Journal of neurotrauma* **32**, 1599-1607 (2015).
158. Alder, J. et al. Genetic and pharmacological intervention of the p75NTR pathway alters morphological and behavioural recovery following traumatic brain injury in mice. *Brain Inj*, 1-18 (2015).

159. Hergenroeder, G. et al. Identification of serum biomarkers in brain-injured adults: potential for predicting elevated intracranial pressure. *J Neurotrauma* **25**, 79-93 (2008).
160. Jeon, S. et al. Role of lipocalin-2-chemokine axis in the development of neuropathic pain following peripheral nerve injury. *The Journal of biological chemistry* **288**, 24116-24127 (2013).
161. Werner, J.K. & Stevens, R.D. Traumatic brain injury: recent advances in plasticity and regeneration. *Curr Opin Neurol* **28**, 565-573 (2015).
162. Escher, C., Reiter, L., Ossola, R., MacLean, B. and Rinner O. in American Association for Mass Spectrometry (Vancouver, BC, Canada; 2012).
163. Medicine, U.S.N.L.o. (U.S. National Library of Medicine, 2016).
164. Rose, C., Parker, A., Jefferson, B. & Cartmell, E. The Characterization of Feces and Urine: A Review of the Literature to Inform Advanced Treatment Technology. *Critical Reviews in Environmental Science and Technology* **45**, 1827-1879 (2015).
165. Kim, M.J., Frankel, A.H. & Tam, F.W. Urine proteomics and biomarkers in renal disease. *Nephron Exp Nephrol* **119**, e1-7 (2011).
166. Kim, S.C., Page, E.K. & Knechtle, S.J. Urine proteomics in kidney transplantation. *Transplant Rev (Orlando)* **28**, 15-20 (2014).
167. Magalhaes, P., Mischak, H. & Zurbig, P. Urinary proteomics using capillary electrophoresis coupled to mass spectrometry for diagnosis and prognosis in kidney diseases. *Curr Opin Nephrol Hypertens* **25**, 494-501 (2016).
168. Knepper, M.A. & Pisitkun, T. Exosomes in urine: who would have thought...? *Kidney Int* **72**, 1043-1045 (2007).
169. Alvarez, M.L. Isolation of urinary exosomes for RNA biomarker discovery using a simple, fast, and highly scalable method. *Methods Mol Biol* **1182**, 145-170 (2014).
170. Gonzales, P., Pisitkun, T. & Knepper, M.A. Urinary exosomes: is there a future? *Nephrol Dial Transplant* **23**, 1799-1801 (2008).
171. Zhou, H. et al. Collection, storage, preservation, and normalization of human urinary exosomes for biomarker discovery. *Kidney Int* **69**, 1471-1476 (2006).
172. Dear, J.W., Street, J.M. & Bailey, M.A. Urinary exosomes: a reservoir for biomarker discovery and potential mediators of intrarenal signalling. *Proteomics* **13**, 1572-1580 (2013).
173. Moon, P.G., You, S., Lee, J.E., Hwang, D. & Baek, M.C. Urinary exosomes and proteomics. *Mass Spectrom Rev* **30**, 1185-1202 (2011).
174. Polpitiya, A.D. et al. DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* **24**, 1556-1558 (2008).
175. Fernandez-Llama, P. et al. Tamm-Horsfall protein and urinary exosome isolation. *Kidney Int* **77**, 736-742 (2010).
176. Gonzales, P.A. et al. Large-scale proteomics and phosphoproteomics of urinary exosomes. *J Am Soc Nephrol* **20**, 363-379 (2009).

177. Gonzales, P.A. et al. Isolation and purification of exosomes in urine. *Methods Mol Biol* **641**, 89-99 (2010).
178. Pathan, M. et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics* **15**, 2597-2601 (2015).
179. Keerthikumar, S. et al. ExoCarta: A Web-Based Compendium of Exosomal Cargo. *J Mol Biol* **428**, 688-692 (2016).
180. Mathivanan, S., Fahner, C.J., Reid, G.E. & Simpson, R.J. ExoCarta 2012: database of exosomal proteins, RNA and lipids. *Nucleic Acids Res* **40**, D1241-1244 (2012).
181. Mathivanan, S. & Simpson, R.J. ExoCarta: A compendium of exosomal proteins and RNA. *Proteomics* **9**, 4997-5000 (2009).
182. Taylor, S.L., Ruhaak, L.R., Kelly, K., Weiss, R.H. & Kim, K. Effects of imputation on correlation: implications for analysis of mass spectrometry data from multiple biological matrices. *Brief Bioinform* (2016).
183. Richens, J.L. et al. Practical detection of a definitive biomarker panel for Alzheimer's disease; comparisons between matched plasma and cerebrospinal fluid. *Int J Mol Epidemiol Genet* **5**, 53-70 (2014).
184. Begaz, T., Kyriacou, D.N., Segal, J. & Bazarian, J.J. Serum biochemical markers for post-concussion syndrome in patients with mild traumatic brain injury. *J Neurotrauma* **23**, 1201-1210 (2006).
185. de Bousard, C.N. et al. S100 and cognitive impairment after mild traumatic brain injury. *J Rehabil Med* **37**, 53-57 (2005).
186. Kleinert, K., Schleich, F., Biasca, N. & Simmen, H.P. [Is there a correlation between S100 beta and post-concussion symptoms after mild traumatic brain injury?]. *Zentralbl Chir* **135**, 277-278 (2010).
187. Eikmans, M. et al. Expression of surfactant protein-C, S100A8, S100A9, and B cell markers in renal allografts: investigation of the prognostic value. *J Am Soc Nephrol* **16**, 3771-3786 (2005).
188. Fujiu, K., Manabe, I. & Nagai, R. Renal collecting duct epithelial cells regulate inflammation in tubulointerstitial damage in mice. *J Clin Invest* **121**, 3425-3441 (2011).
189. Zhang, L. et al. Proteins S100A8 and S100A9 are potential biomarkers for renal cell carcinoma in the early stages: results from a proteomic study integrated with bioinformatics analysis. *Mol Med Rep* **11**, 4093-4100 (2015).
190. Suzuki, T. et al. Molecular cloning of a novel apoptosis-related gene, human Nap1 (NCKAP1), and its possible relation to Alzheimer disease. *Genomics* **63**, 246-254 (2000).
191. Yamamoto, A., Suzuki, T. & Sakaki, Y. Isolation of hNap1BP which interacts with human Nap1 (NCKAP1) whose expression is down-regulated in Alzheimer's disease. *Gene* **271**, 159-169 (2001).
192. Mattson, M.P. Apoptosis in neurodegenerative disorders. *Nat Rev Mol Cell Biol* **1**, 120-129 (2000).

193. Jackson, R.S., 2nd, Cho, Y.J., Stein, S. & Liang, P. CYFIP2, a direct p53 target, is leptomycin-B sensitive. *Cell Cycle* **6**, 95-103 (2007).
194. Mayne, M. et al. CYFIP2 is highly abundant in CD4+ cells from multiple sclerosis patients and is involved in T cell adhesion. *Eur J Immunol* **34**, 1217-1227 (2004).
195. Saller, E. et al. Increased apoptosis induction by 121F mutant p53. *EMBO J* **18**, 4424-4437 (1999).
196. Schenck, A., Bardoni, B., Moro, A., Bagni, C. & Mandel, J.L. A highly conserved protein family interacting with the fragile X mental retardation protein (FMRP) and displaying selective interactions with FMRP-related proteins FXR1P and FXR2P. *Proc Natl Acad Sci U S A* **98**, 8844-8849 (2001).
197. Hunter, D.J. et al. A pathway and approach to biomarker validation and qualification for osteoarthritis clinical trials. *Curr Drug Targets* **11**, 536-545 (2010).
198. Caceres, J.A. & Goldstein, J.N. Intracranial hemorrhage. *Emerg Med Clin North Am* **30**, 771-794 (2012).
199. Kushner, D. Mild traumatic brain injury: toward understanding manifestations and treatment. *Arch Intern Med* **158**, 1617-1624 (1998).
200. van Gijn, J., Kerr, R.S. & Rinkel, G.J. Subarachnoid haemorrhage. *Lancet* **369**, 306-318 (2007).
201. Dorayappan, K.D., Wallbillich, J.J., Cohn, D.E. & Selvendiran, K. The biological significance and clinical applications of exosomes in ovarian cancer. *Gynecol Oncol* **142**, 199-205 (2016).
202. Lin, J. et al. Exosomes: novel biomarkers for clinical diagnosis. *ScientificWorldJournal* **2015**, 657086 (2015).
203. Finlay, B.B. & McFadden, G. Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. *Cell* **124**, 767-782 (2006).
204. Hornef, M.W., Wick, M.J., Rhen, M. & Normark, S. Bacterial strategies for overcoming host innate and adaptive immune responses. *Nat Immunol* **3**, 1033-1040 (2002).
205. Li, K. et al. Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proc Natl Acad Sci U S A* **102**, 2992-2997 (2005).
206. Li, X.D., Sun, L., Seth, R.B., Pineda, G. & Chen, Z.J. Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. *Proc Natl Acad Sci U S A* **102**, 17717-17722 (2005).
207. Navarro, L., Alto, N.M. & Dixon, J.E. Functions of the Yersinia effector proteins in inhibiting host immune responses. *Curr Opin Microbiol* **8**, 21-27 (2005).
208. Park, J.M., Greten, F.R., Li, Z.W. & Karin, M. Macrophage apoptosis by anthrax lethal factor through p38 MAP kinase inhibition. *Science* **297**, 2048-2051 (2002).
209. McCann, H., Stevens, C.H., Cartwright, H. & Halliday, G.M. alpha-Synucleinopathy phenotypes. *Parkinsonism Relat Disord* **20 Suppl 1**, S62-67 (2014).

210. Peelaerts, W. et al. alpha-Synuclein strains cause distinct synucleinopathies after local and systemic administration. *Nature* **522**, 340-344 (2015).
211. Mason, S.D. & Joyce, J.A. Proteolytic networks in cancer. *Trends Cell Biol* **21**, 228-237 (2011).
212. Sevenich, L. & Joyce, J.A. Pericellular proteolysis in cancer. *Genes Dev* **28**, 2331-2347 (2014).
213. Decramer, S. et al. Urine in clinical proteomics. *Mol Cell Proteomics* **7**, 1850-1862 (2008).
214. Zheng, J., Liu, L., Wang, J. & Jin, Q. Urinary proteomic and non-prefractionation quantitative phosphoproteomic analysis during pregnancy and non-pregnancy. *BMC Genomics* **14**, 777 (2013).
215. Doucet, A., Butler, G.S., Rodriguez, D., Prudova, A. & Overall, C.M. Metadegradomics: toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. *Mol Cell Proteomics* **7**, 1925-1951 (2008).
216. Alves, P. et al. Fast and accurate identification of semi-tryptic peptides in shotgun proteomics. *Bioinformatics* **24**, 102-109 (2008).
217. Rodriguez, J., Gupta, N., Smith, R.D. & Pevzner, P.A. Does trypsin cut before proline? *J Proteome Res* **7**, 300-305 (2008).
218. Brown, R.C., Lockwood, A.H. & Sonawane, B.R. Neurodegenerative diseases: an overview of environmental risk factors. *Environ Health Perspect* **113**, 1250-1256 (2005).
219. Hung, C.W., Chen, Y.C., Hsieh, W.L., Chiou, S.H. & Kao, C.L. Ageing and neurodegenerative diseases. *Ageing Res Rev* **9 Suppl 1**, S36-46 (2010).
220. Zoghbi, H.Y. & Orr, H.T. Pathogenic mechanisms of a polyglutamine-mediated neurodegenerative disease, spinocerebellar ataxia type 1. *The Journal of biological chemistry* **284**, 7425-7429 (2009).
221. DiMauro, S. & Schon, E.A. Mitochondrial disorders in the nervous system. *Annu Rev Neurosci* **31**, 91-123 (2008).
222. Lin, M.T. & Beal, M.F. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* **443**, 787-795 (2006).
223. Vila, M. & Przedborski, S. Targeting programmed cell death in neurodegenerative diseases. *Nat Rev Neurosci* **4**, 365-375 (2003).
224. Henley, S.M., Bates, G.P. & Tabrizi, S.J. Biomarkers for neurodegenerative diseases. *Curr Opin Neurol* **18**, 698-705 (2005).
225. Blumenfeld, H. Neuroanatomy through clinical cases, Edn. 2nd. (Sinauer Associates, Sunderland, Mass.; 2010).
226. Gulbrandsen, A. et al. In-depth characterization of the cerebrospinal fluid (CSF) proteome displayed through the CSF proteome resource (CSF-PR). *Mol Cell Proteomics* **13**, 3152-3163 (2014).
227. Beck, H.C., Overgaard, M. & Melholt Rasmussen, L. Plasma proteomics to identify biomarkers – application to cardiovascular diseases. *Translational Proteomics* **7**, 40-48 (2015).

-
228. De Bock, M. et al. Challenges for biomarker discovery in body fluids using SELDI-TOF-MS. *J Biomed Biotechnol* **2010**, 906082 (2010).
229. Million, R. et al. High abundance proteins depletion vs low abundance proteins enrichment: comparison of methods to reduce the plasma proteome complexity. *PLoS one* **6**, e19603 (2011).
230. Parker, C.E. & Borchers, C.H. Mass spectrometry based biomarker discovery, verification, and validation--quality assurance and control of protein biomarker assays. *Mol Oncol* **8**, 840-858 (2014).
231. Ramstrom, M. et al. Development of affinity columns for the removal of high-abundance proteins in cerebrospinal fluid. *Biotechnol Appl Biochem* **52**, 159-166 (2009).
232. Padmanabhan, J., Levy, M., Dickson, D.W. & Potter, H. Alpha1-antichymotrypsin, an inflammatory protein overexpressed in Alzheimer's disease brain, induces tau phosphorylation in neurons. *Brain* **129**, 3020-3034 (2006).
233. Tyagi, E., Fiorelli, T., Norden, M. & Padmanabhan, J. Alpha 1-Antichymotrypsin, an Inflammatory Protein Overexpressed in the Brains of Patients with Alzheimer's Disease, Induces Tau Hyperphosphorylation through c-Jun N-Terminal Kinase Activation. *Int J Alzheimers Dis* **2013**, 606083 (2013).
234. Kalsheker, N.A. Alpha 1-antichymotrypsin. *Int J Biochem Cell Biol* **28**, 961-964 (1996).
235. Anoop, A., Singh, P.K., Jacob, R.S. & Maji, S.K. CSF Biomarkers for Alzheimer's Disease Diagnosis. *Int J Alzheimers Dis* **2010** (2010).
236. Wormser, U. et al. Reduced levels of alpha-1-antitrypsin in cerebrospinal fluid of amyotrophic lateral sclerosis patients: a novel approach for a potential treatment. *J Neuroinflammation* **13**, 131 (2016).
237. Kolev, M.V., Ruseva, M.M., Harris, C.L., Morgan, B.P. & Donev, R.M. Implication of complement system and its regulators in Alzheimer's disease. *Curr Neuropharmacol* **7**, 1-8 (2009).
238. Orsini, F., De Blasio, D., Zangari, R., Zanier, E.R. & De Simoni, M.G. Versatility of the complement system in neuroinflammation, neurodegeneration and brain homeostasis. *Front Cell Neurosci* **8**, 380 (2014).
239. Shen, Y., Yang, L. & Li, R. What does complement do in Alzheimer's disease? Old molecules with new insights. *Transl Neurodegener* **2**, 21 (2013).
240. Wang, B., Matsuoka, S., Carpenter, P.B. & Elledge, S.J. 53BP1, a mediator of the DNA damage checkpoint. *Science* **298**, 1435-1438 (2002).
241. Blackwell, L.J., Bjornson, K.P. & Modrich, P. DNA-dependent activation of the hMutSalpha ATPase. *The Journal of biological chemistry* **273**, 32049-32054 (1998).
242. Blackwell, L.J., Martik, D., Bjornson, K.P., Bjornson, E.S. & Modrich, P. Nucleotide-promoted release of hMutSalpha from heteroduplex DNA is consistent with an ATP-dependent translocation mechanism. *The Journal of biological chemistry* **273**, 32055-32062 (1998).
243. Mayeux, R. Biomarkers: potential uses and limitations. *NeuroRx* **1**, 182-188 (2004).

-
244. Yuce, O. & West, S.C. Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. *Mol Cell Biol* **33**, 406-417 (2013).
245. Kulkarni, A. & Wilson, D.M., 3rd The involvement of DNA-damage and -repair defects in neurological dysfunction. *Am J Hum Genet* **82**, 539-566 (2008).
246. Gollin, P.A., Kalaria, R.N., Eikelenboom, P., Rozemuller, A. & Perry, G. Alpha 1-antitrypsin and alpha 1-antichymotrypsin are in the lesions of Alzheimer's disease. *Neuroreport* **3**, 201-203 (1992).
247. Lieberman, J., Schleissner, L., Tachiki, K.H. & Kling, A.S. Serum alpha 1-antichymotrypsin level as a marker for Alzheimer-type dementia. *Neurobiol Aging* **16**, 747-753 (1995).
248. Fukushima, N., Furuta, D., Hidaka, Y., Moriyama, R. & Tsujiuchi, T. Post-translational modifications of tubulin in the nervous system. *J Neurochem* **109**, 683-693 (2009).
249. Ikegami, K. et al. TTL7 is a mammalian beta-tubulin polyglutamylase required for growth of MAP2-positive neurites. *The Journal of biological chemistry* **281**, 30707-30716 (2006).
250. Zhang, F. et al. Posttranslational modifications of alpha-tubulin in Alzheimer disease. *Transl Neurodegener* **4**, 9 (2015).
251. Geroldi, D. et al. Genetic association of alpha2-Heremans-Schmid glycoprotein polymorphism with late-onset Alzheimer's disease in Italians. *Neurosci Lett* **386**, 176-178 (2005).
252. Heit, C. et al. Update of the human and mouse SERPIN gene superfamily. *Hum Genomics* **7**, 22 (2013).
253. Silverman, G.A. et al. The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *The Journal of biological chemistry* **276**, 33293-33296 (2001).
254. Han, M.H. et al. Proteomic analysis of active multiple sclerosis lesions reveals therapeutic targets. *Nature* **451**, 1076-1081 (2008).
255. Gatto, M. et al. Serpins, immunity and autoimmunity: old molecules, new functions. *Clin Rev Allergy Immunol* **45**, 267-280 (2013).
256. Aspelund, A. et al. A dural lymphatic vascular system that drains brain interstitial fluid and macromolecules. *J Exp Med* **212**, 991-999 (2015).
257. Yin, X. et al. Glycoproteomic analysis of the secretome of human endothelial cells. *Mol Cell Proteomics* **12**, 956-978 (2013).
258. O'Keefe, F.M. et al. Loss of insight in frontotemporal dementia, corticobasal degeneration and progressive supranuclear palsy. *Brain* **130**, 753-764 (2007).
259. Wadia, P.M. & Lang, A.E. The many faces of corticobasal degeneration. *Parkinsonism Relat Disord* **13 Suppl 3**, S336-340 (2007).
260. Nielsen, H.M. et al. Plasma and CSF serpins in Alzheimer disease and dementia with Lewy bodies. *Neurology* **69**, 1569-1579 (2007).

261. Halbgebauer, S. et al. Modified serpinA1 as risk marker for Parkinson's disease dementia: Analysis of baseline data. *Sci Rep* **6**, 26145 (2016).
262. Skates, S.J. et al. Statistical design for biospecimen cohort size in proteomics-based biomarker discovery and verification studies. *J Proteome Res* **12**, 5383-5394 (2013).

(blank page)

8. Appendix

8.1 MStern Blot

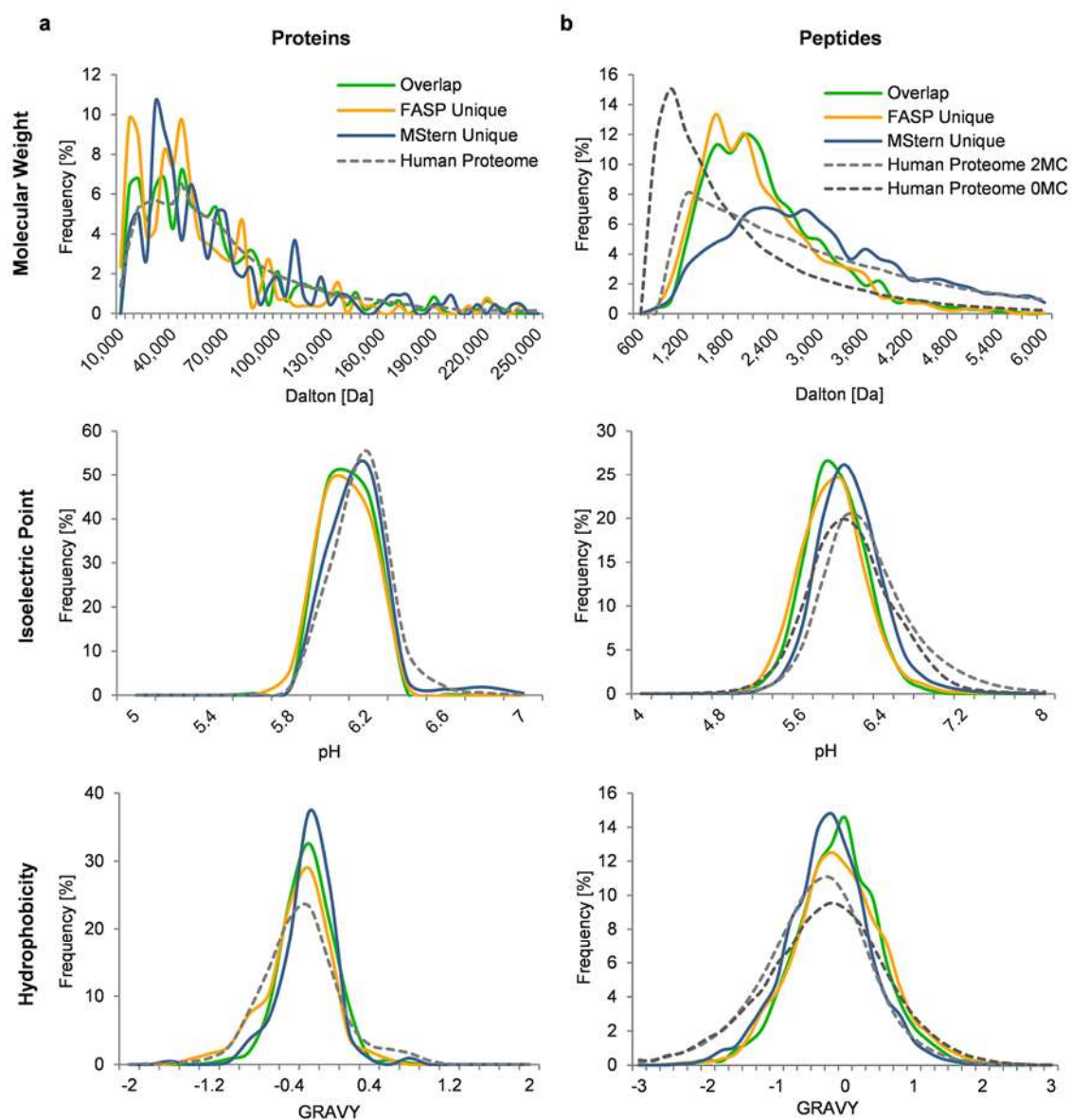


Figure 8.1.1: Physico-chemical Properties (CSF). **a)** Comparison of three different properties: Molecular Weight (top), isoelectric pH (middle) and GRAVY score (bottom) at protein level for MStern blotting-specific proteins (blue trace), FASP-specific proteins (yellow trace), shared proteins (green trace) and theoretical distribution of the entire human proteome (dashed grey trace). **b)** Comparison of physico-chemical property changes at peptide level: Molecular Weight (top), isoelectric pH (middle) and GRAVY score (bottom) for MStern blotting-specific peptides (blue trace), FASP-specific peptides (yellow trace), shared peptides (green trace) and theoretical distribution upon tryptically digesting the entire human proteome assuming no missed cleavages (0 MC; dashed dark grey trace) or 2 missed cleavages (2 MC; dashed light grey trace).

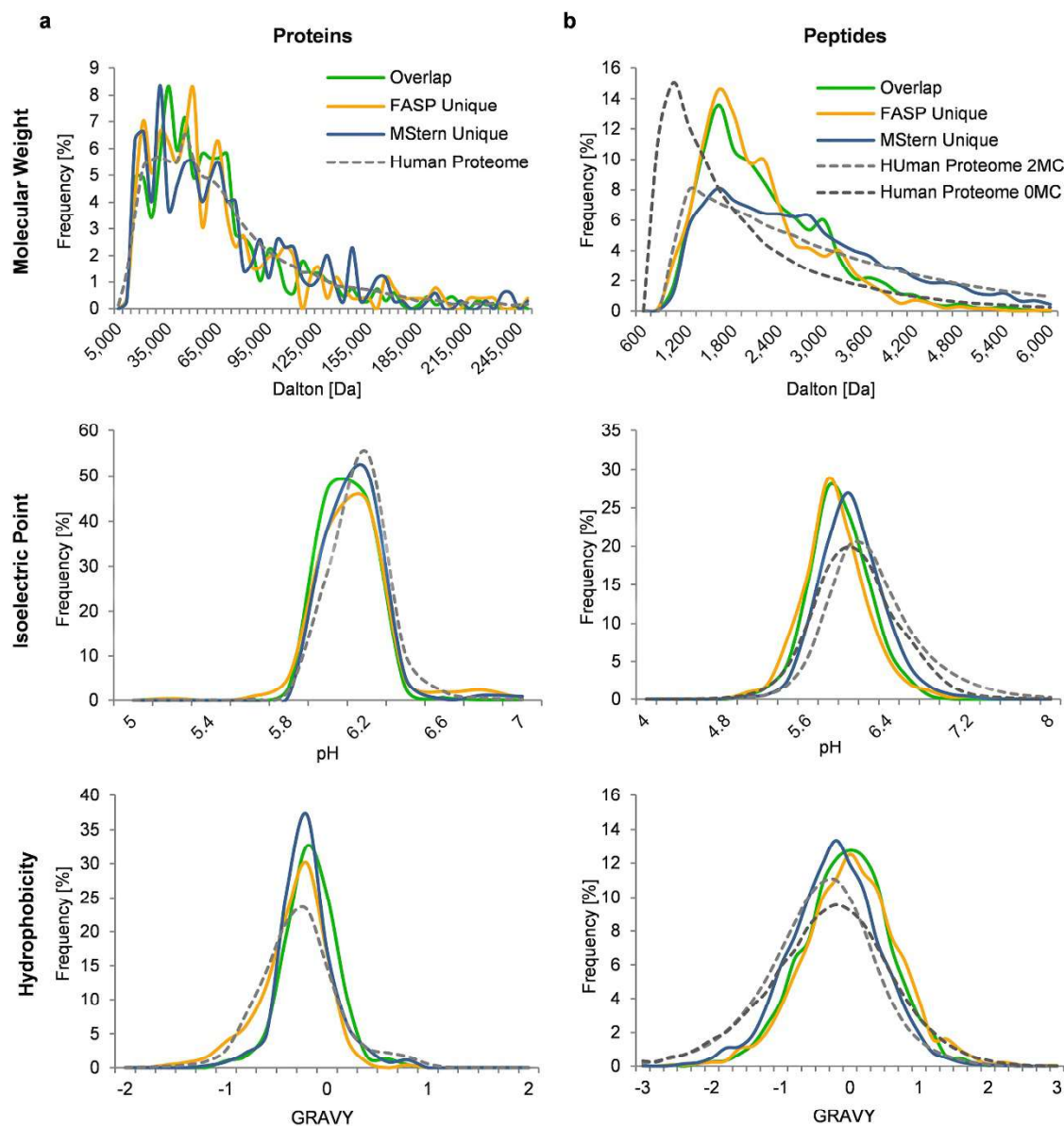


Figure 8.1.2: Physico-chemical Properties (Urine). **a**) Comparison of three different properties: Molecular Weight (top), isoelectric pH (middle) and GRAVY score (bottom) at protein level for MStern blotting-specific proteins (blue trace), FASP-specific proteins (yellow trace), shared proteins (green trace) and theoretical distribution of the entire human proteome (dashed grey trace). **b**) Comparison of physico-chemical property changes at peptide level: Molecular Weight (top), isoelectric pH (middle) and GRAVY score (bottom) for MStern blotting-specific peptides (blue trace), FASP-specific peptides (yellow trace), shared peptides (green trace) and theoretical distribution upon tryptically digesting the entire human proteome assuming no missed cleavages (0 MC; dashed dark grey trace) or 2 missed cleavages (2 MC; dashed light grey trace).

8.2 Urinary Biomarker Discovery for mTBI

8.2.1 PCA Plots of SWATH-based Quantification

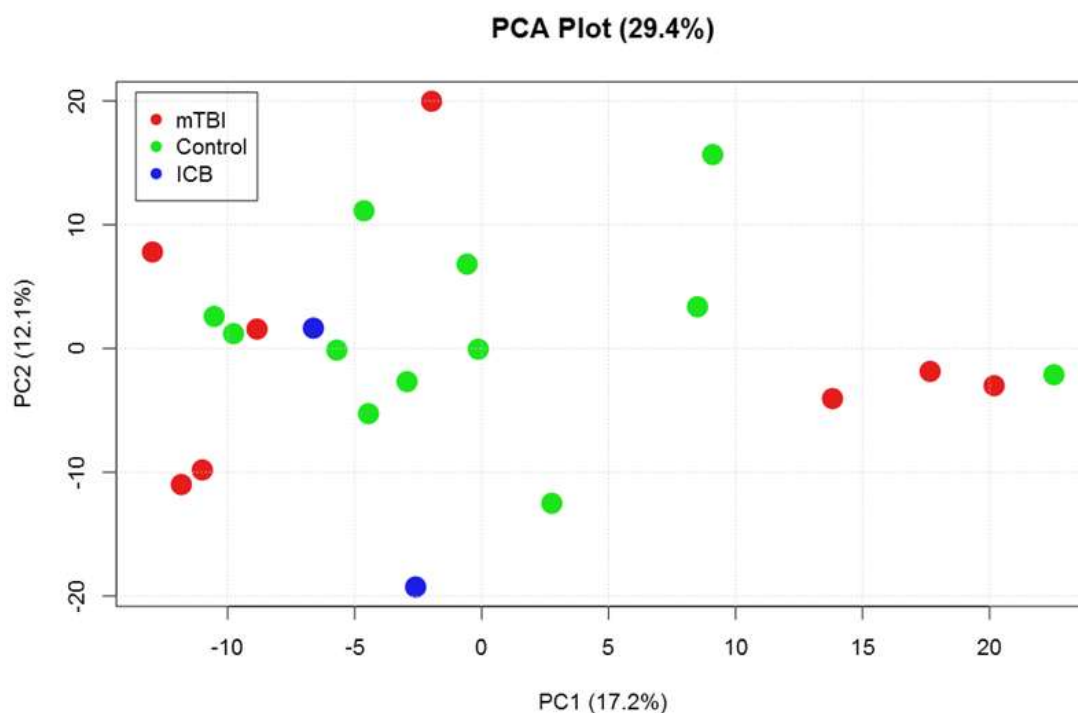


Figure 8.2.1: Unsupervised mTBI PCA Plot (1). All quantified proteins entered for PCA analysis. Plot accounts for a total of 29.4 % of the dataset variance. PC 1 (17.2 %; X – axis) and PC 2 (12.1 %; Y – axis) are shown. No clear separation of clinical groups can be deviated. Clinical groups defined as: mTBI (red), Controls (green) and ICB (blue).

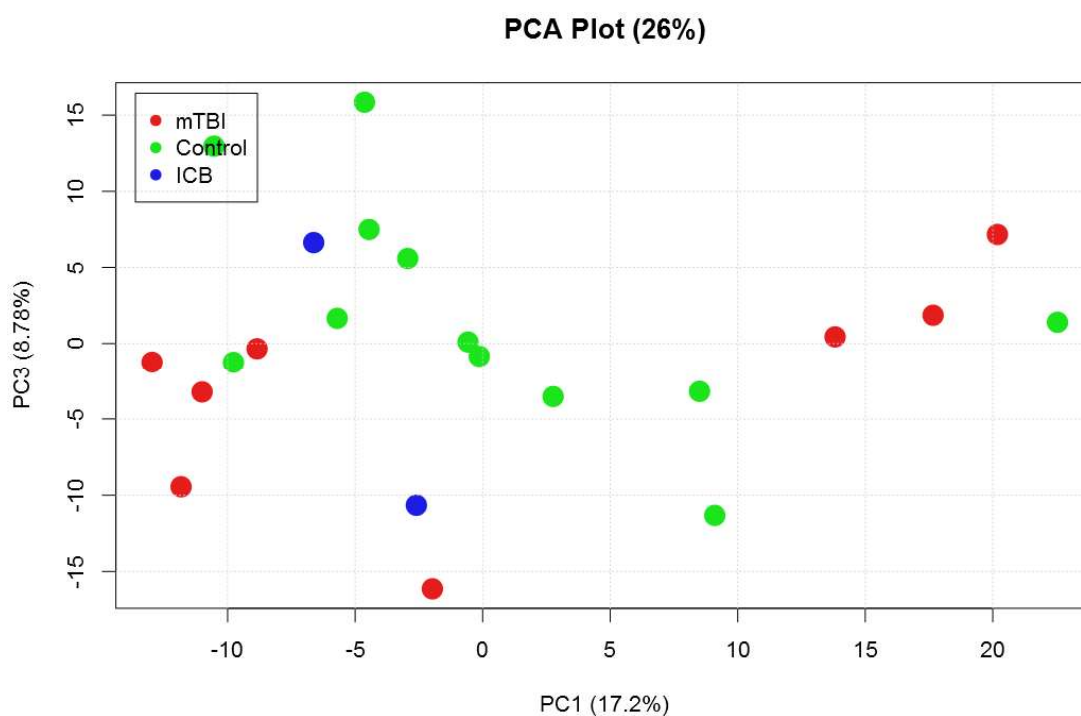


Figure 8.2.2: Unsupervised mTBI PCA Plot (2). All quantified proteins entered for PCA analysis. Plot accounts for a total of 26 % of the dataset variance. PC 1 (17.2 %; X – axis) and PC 3 (8.8 %; Y – axis) are shown. No clear separation of clinical groups can be deviated. Clinical groups defined as: mTBI (red), Controls (green) and ICB (blue).

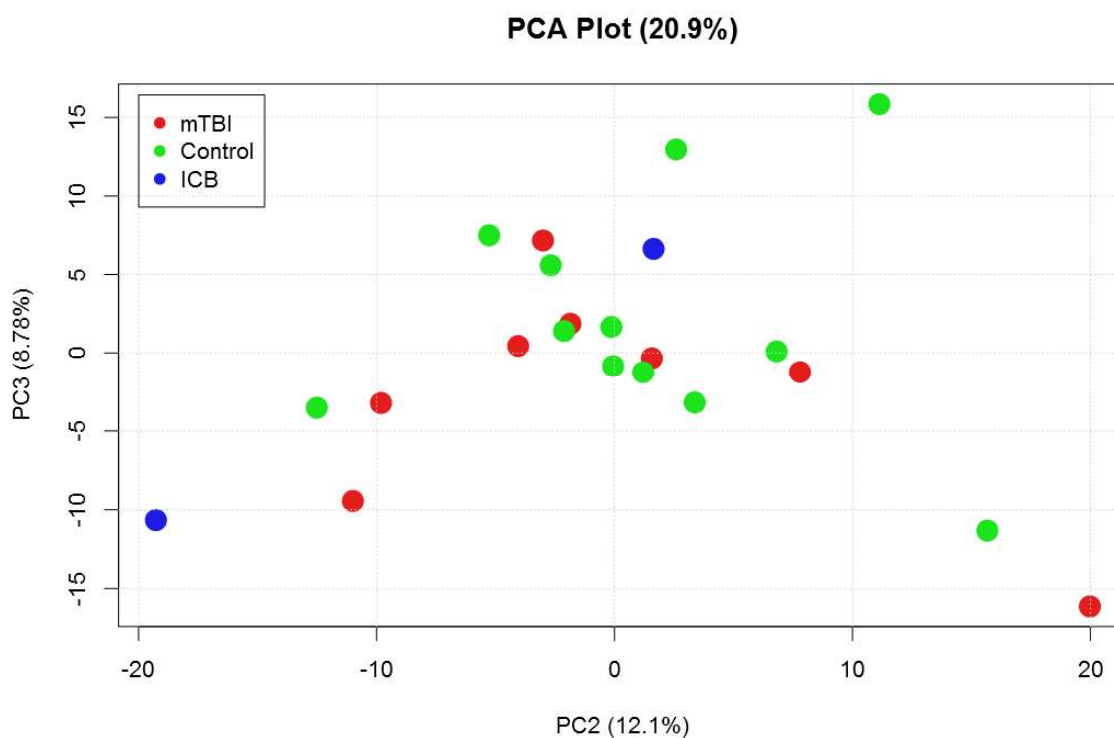


Figure 8.2.3: Unsupervised mTBI PCA Plot (3). All quantified proteins entered for PCA analysis. Plot accounts for a total of ~ 21 % of the dataset variance. PC 2 (12.1 % - X - axis) and PC 3 (8.8 % - Y - axis) are shown. No clear separation of clinical groups can be deviated. Clinical groups defined as: mTBI (red), Controls (green) and ICB (blue).

8.2.2 PCA Plots of the secondary discovery study

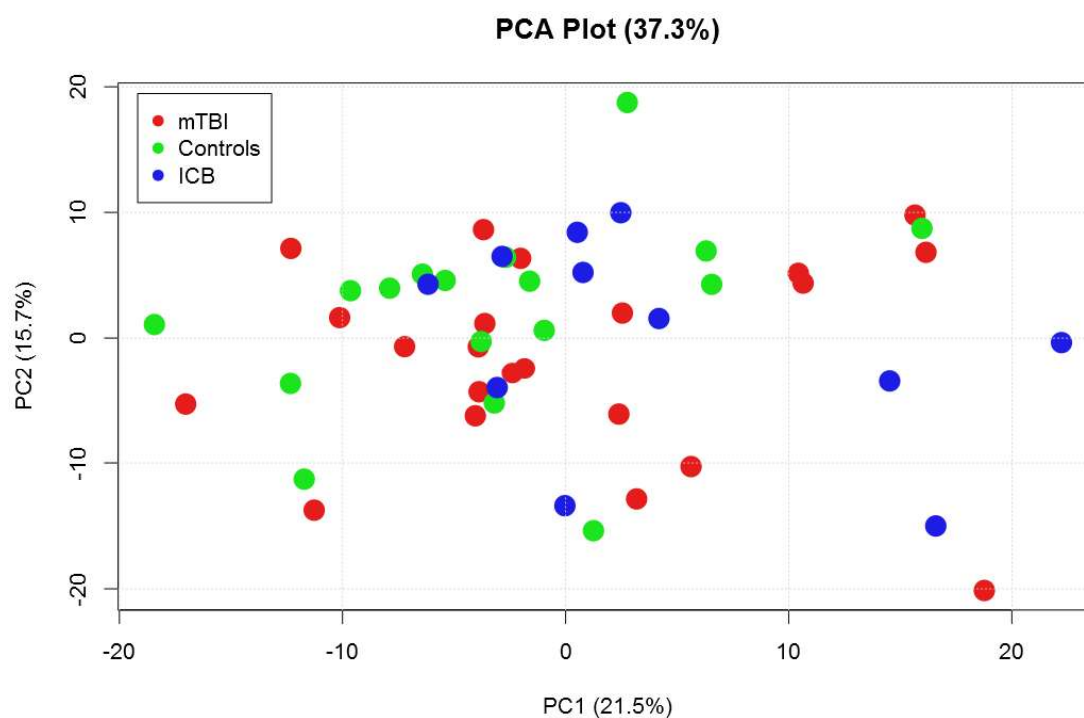


Figure 8.2.4: Unsupervised mTBI PCA Plot for secondary discovery (1). All quantified proteins entered for PCA analysis. Plot accounts for a total of 37.3 % of the dataset variance. PC 1 (21.5 % - X - axis) and PC 2 (15.7 % - Y - axis) are shown. No clear separation of clinical groups can be deviated. Clinical groups defined as: mTBI (red), Controls (green) and ICB (blue).

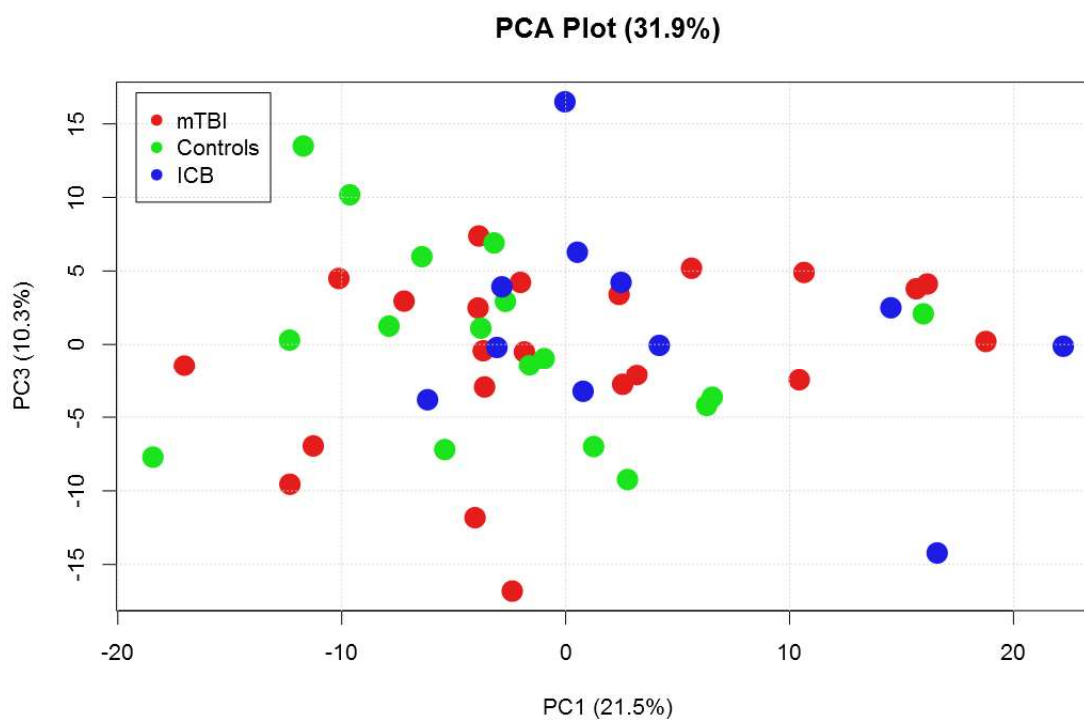


Figure 8.2.5: Unsupervised mTBI PCA Plot for secondary discovery (2). All quantified proteins entered for PCA analysis. Plot accounts for a total of ~32 % of the dataset variance. PC 1 (21.5 % - X - axis) and PC 2 (10.3 % - Y - axis) are shown. No clear separation of clinical groups can be deviated. Clinical groups defined as: mTBI (red), Controls (green) and ICB (blue).

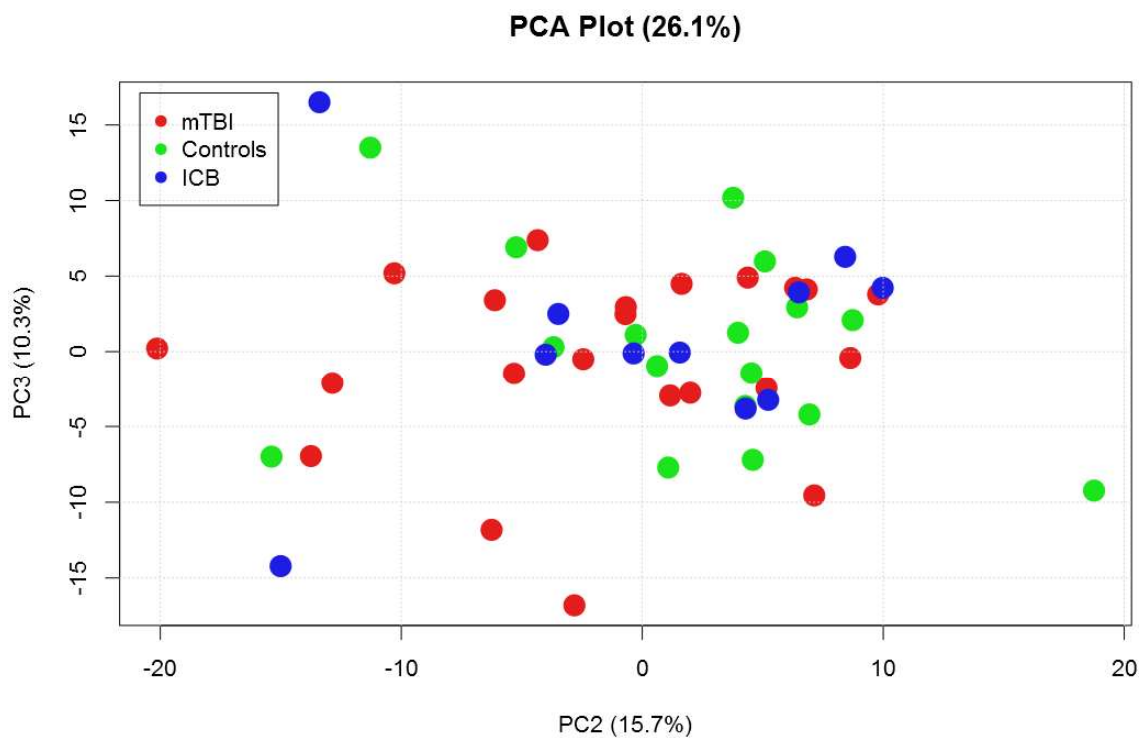


Figure 8.2.6: Unsupervised mTBI PCA Plot for secondary discovery (3). All quantified proteins entered for PCA analysis. Plot accounts for a total of 29.4 % of the dataset variance. PC 2 (15.7 % - X - axis) and PC 3 (10.3 % - Y - axis) are shown. No clear separation of clinical groups can be deviated. Clinical groups defined as: mTBI (red), Controls (green) and ICB (blue).

8.2.3 Prospective mTBI biomarker candidates

Table 8.2.1: List of identified proteins with significant abundance changes (2nd discovery study). Three different p-values based on the type of comparison, are given: ¥ - mTBI samples vs. all control samples (i.e. incl. ICB samples); £ - mTBI vs. negative controls (i.e. excl. ICB samples); β – mTBI samples vs. only ICB samples.

Protein Accession	Gene Code	p-Value [¥]	p-Value [£]	p-Value ^β
P01766 (HV305)	IGHV3-13	0.004598	0.044340	0.004713
P15144 (AMPN)	ANPEP	0.005516	0.013739	0.035692
Q9H1C7 (CYTM1)	CYSTM1	0.006034	0.069892	0.003484
P12821 (ACE)	ACE	0.012362	0.074377	0.013053
O75264 (SIM24)	SMIM24	0.017108	0.054117	0.039186
O75339 (CILP1)	CILP	0.019012	0.533228	0.000182
P29972 (AQP1)	AQP1	0.019012	0.041431	0.066789
P04431 (KV123)	IGKV1-39	0.021651	0.016068	0.214548
P08779 (K1C16)	KRT16	0.022216	0.174001	0.008412
P21926 (CD9)	CD9	0.023383	0.202489	0.006699
P16444 (DPEP1)	DPEP1	0.024604	0.094721	0.032467
Q96DG6 (CMBL)	CMBL	0.025234	0.008798	0.400814
P01743 (HV102)	IGHV1-46	0.025879	0.089258	0.039186
P22732 (GTR5)	SLC2A5	0.026538	0.169560	0.013053
O43895 (XPP2)	XPNPEP2	0.033146	0.057752	0.108723
P55017 (S12A3)	SLC12A3	0.034793	0.174001	0.021942
Q9Y696 (CLIC4)	CLIC4	0.034793	0.023466	0.284932
P01601 (KV109)	IGKV1D-16	0.036509	0.395513	0.003276
P10768 (ESTD)	ESD	0.038296	0.003532	0.908787
P07195 (LDHB)	LDHB	0.038296	0.007761	0.646756
P04004 (VTNC)	VTN	0.040157	0.112731	0.061306
O00182 (LEG9)	LGALS9	0.046207	0.126167	0.066789
P04216 (THY1)	THY1	0.046207	0.084052	0.117403
P31944 (CASPE)	CASP14	0.046207	0.050675	0.207578
P08473 (NEP)	MME	0.048387	0.165204	0.047050
Q9BXJ7 (AMNLS)	AMN	0.048387	0.072106	0.152114

8.2.4 PCA Plots of the ICB-based biomarker discovery

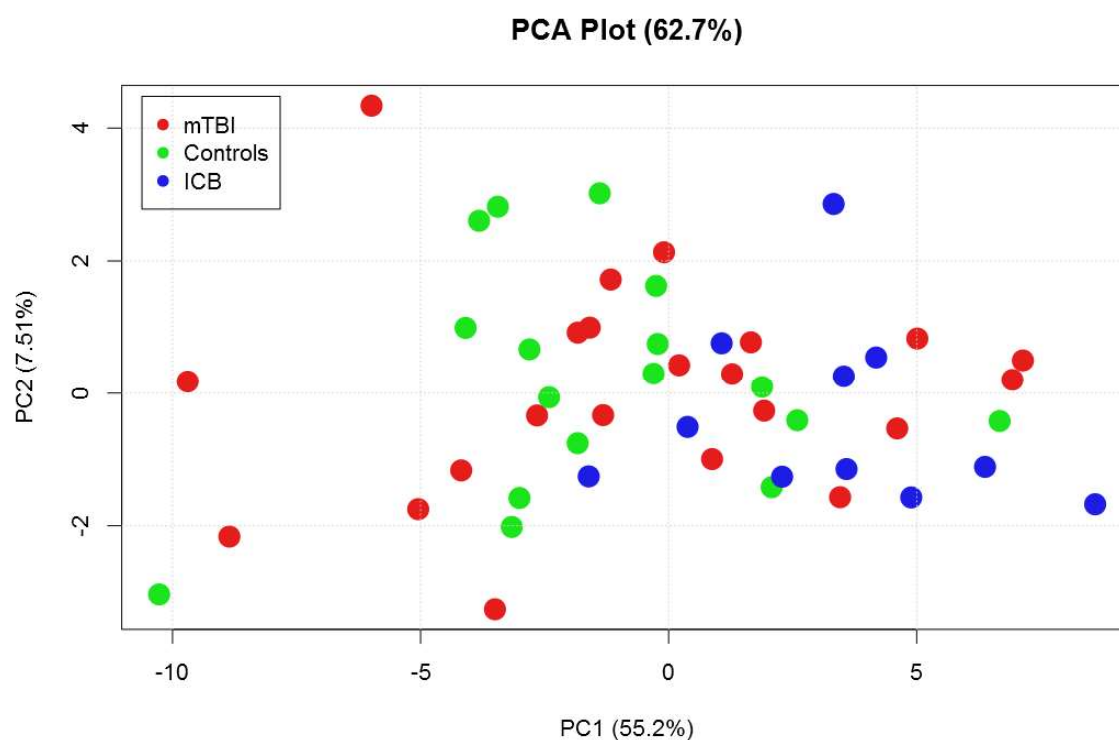


Figure 8.2.7: Unsupervised ICB PCA Plot (1). All quantified proteins entered for PCA analysis. Plot accounts for a total of 29.4 % of the dataset variance. PC 1 (17.2 %; x - axis) and PC 2 (12.1 %; y - axis) are shown. No clear separation of clinical groups can be deviated. Clinical groups defined as: mTBI (red), Controls (green) and ICB (blue).

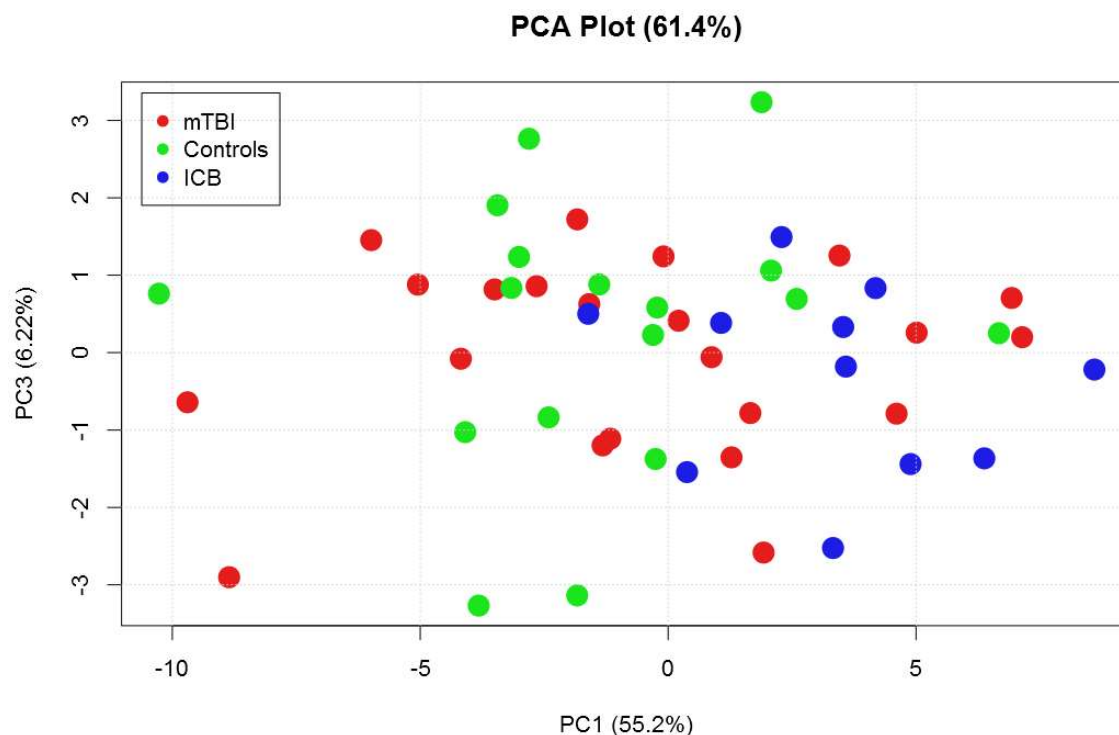


Figure 8.2.8: Unsupervised mTBI PCA Plot (1). All quantified proteins entered for PCA analysis. Plot accounts for a total of 29.4 % of the dataset variance. PC 1 (17.2 %; x - axis) and PC 2 (12.1 %; y - axis) are shown. No clear separation of clinical groups can be deviated. Clinical groups defined as: mTBI (red), Controls (green) and ICB (blue).

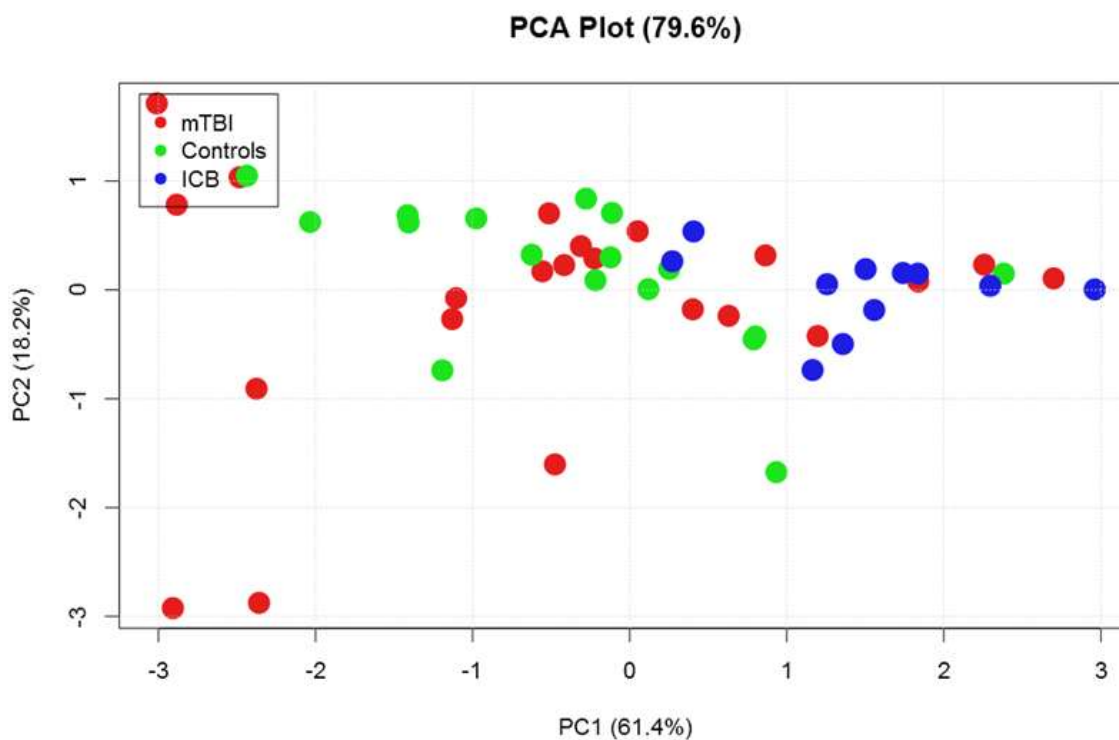


Figure 8.2.9: Supervised PCA plot for ICB biomarker. This plot features only the six most prominent proteins of the AUROC analysis. A total of 79.6 % of the variance (within the 6 analyzed proteins) is displayed. A spread along PC 1 (accounting for 61.4% of variance) is more prominent than the spread of mTBI samples along PC 2 (18.2 %). No clear and no defining clustering tendency can be deviated.

8.3 Exosomal Biomarker Discovery for mTBI

8.3.1 Initial Biomarker Discovery Cohort

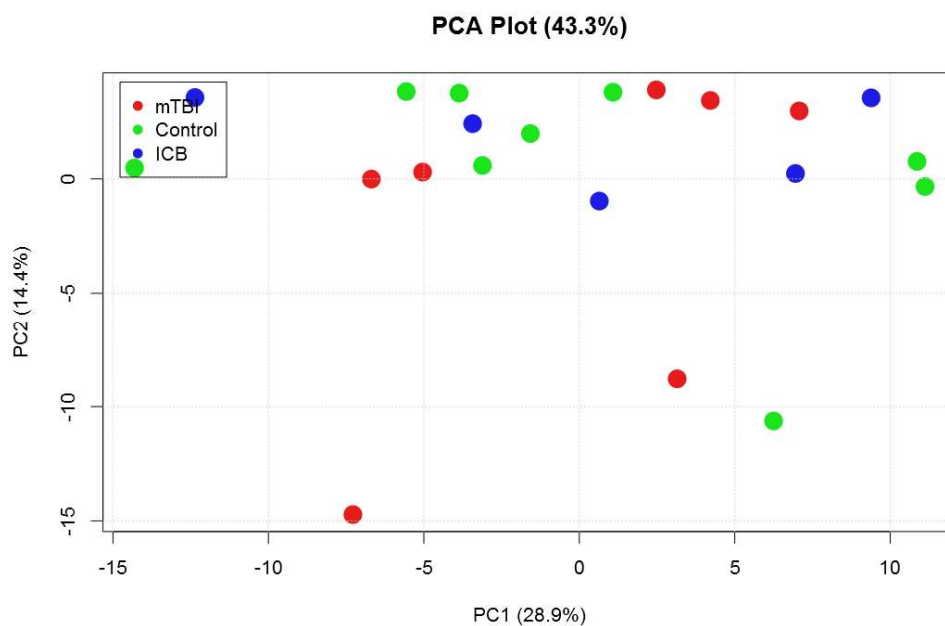


Figure 8.3.1: Unsupervised PCA plot sample clustering. All identified and quantified proteins were entered into this analysis. This plot accounts for a total of 43.3 % variance of the dataset. The first principal component (PC 1) accounts for 28.9% while PC 2 accounts for 14.4 %. Hence, there is a broader spread across PC 1 than PC2. Nonetheless, no group-based separation can be deviated.

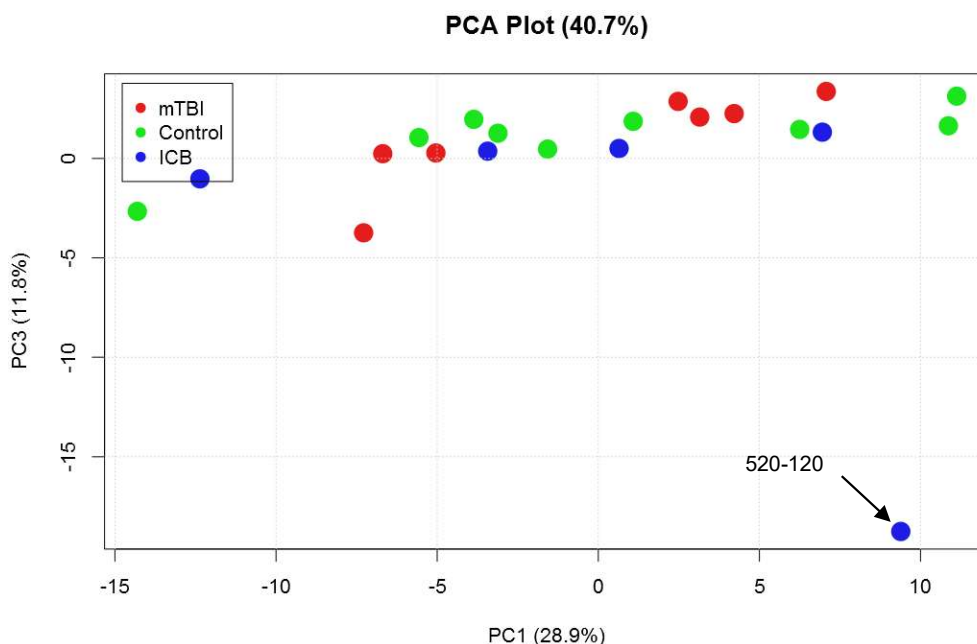


Figure 8.3.2: Unsupervised PCA plot sample clustering. All identified and quantified proteins were entered into this analysis. This plot accounts for a total of 40.7 % variance of the dataset. The first principal component (PC 1) accounts for 28.9% while PC 3 accounts for 11.8 %. Hence, there is a broader spread across PC 1, almost none for PC 3. Massive outlier represented by sample with ID 520-120.

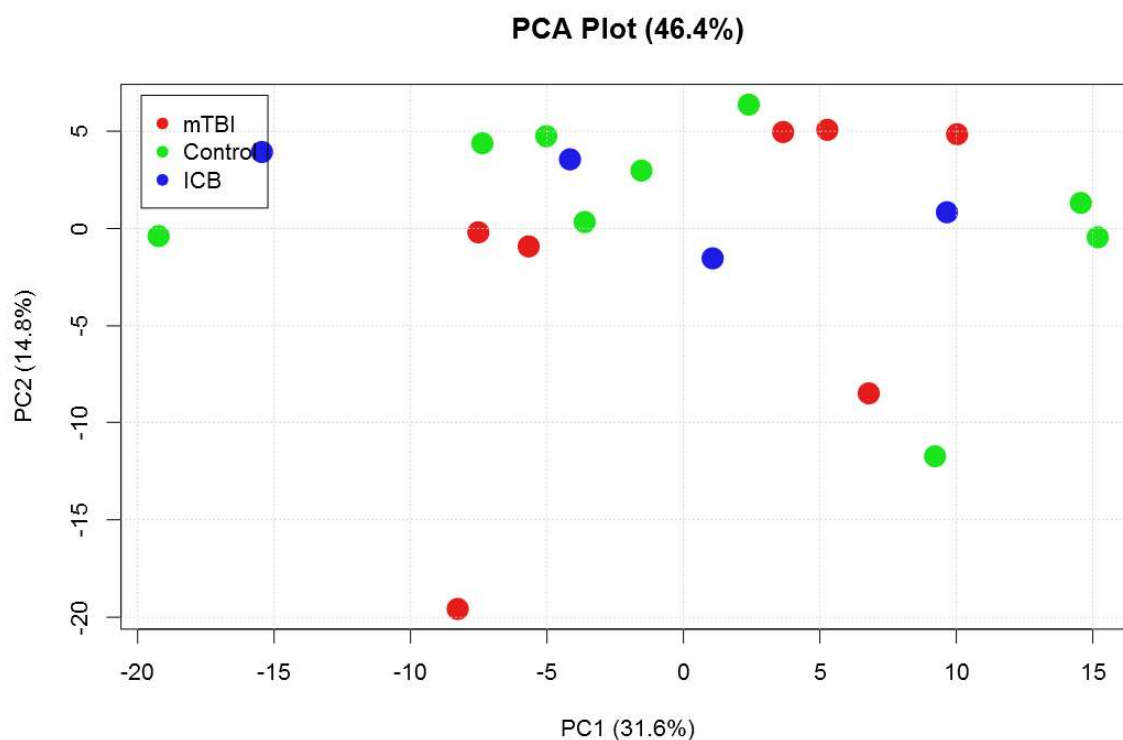


Figure 8.3.3: Semi-Supervised PCA plot. Sample with massive outlier tendencies (520-120) was removed from further consideration. PCA plot accounts for a total of 46.4 % dataset variance with PC 1 of 31.6 % and PC 2 of 14.8 %. The removal of the outlier could not confirm an underlying group-based clustering tendency.

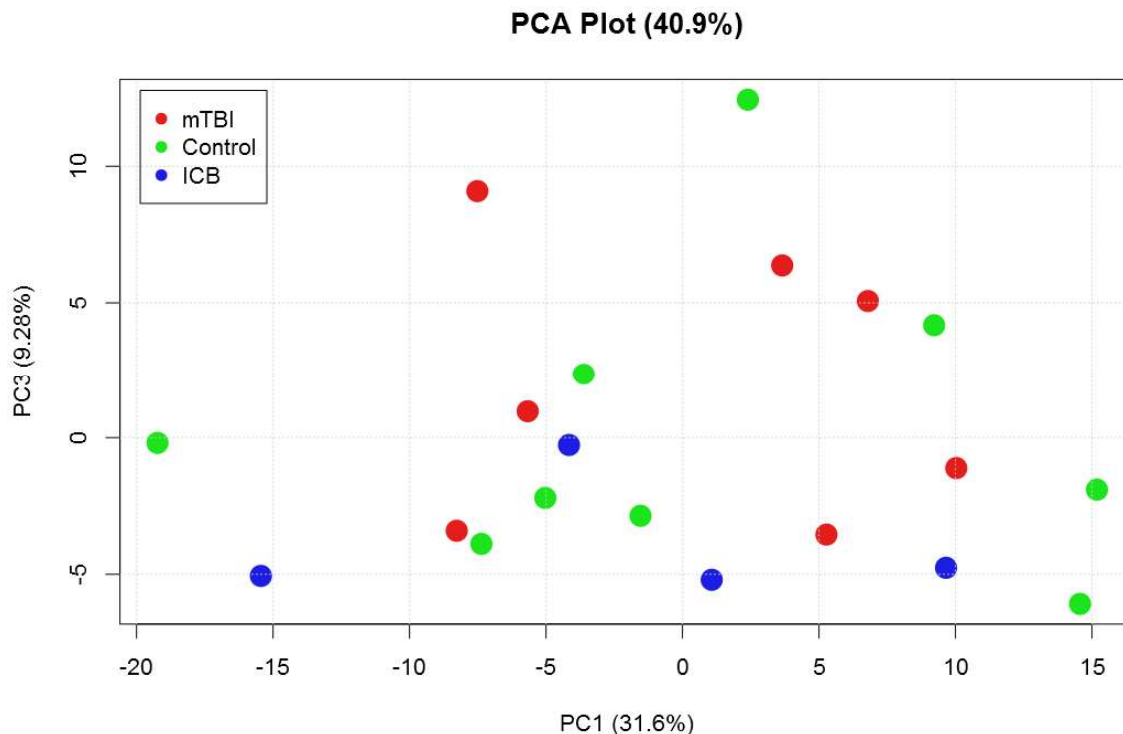


Figure 8.3.4: Semi-Supervised PCA plot. Sample with massive outlier tendencies (520-120) was removed from further consideration. PCA plot accounts for a total of 40.9 % dataset variance with PC 1 of 31.6 % and PC 2 of 9.28 %. The removal of the outlier did not reveal any underlying segregation tendencies.

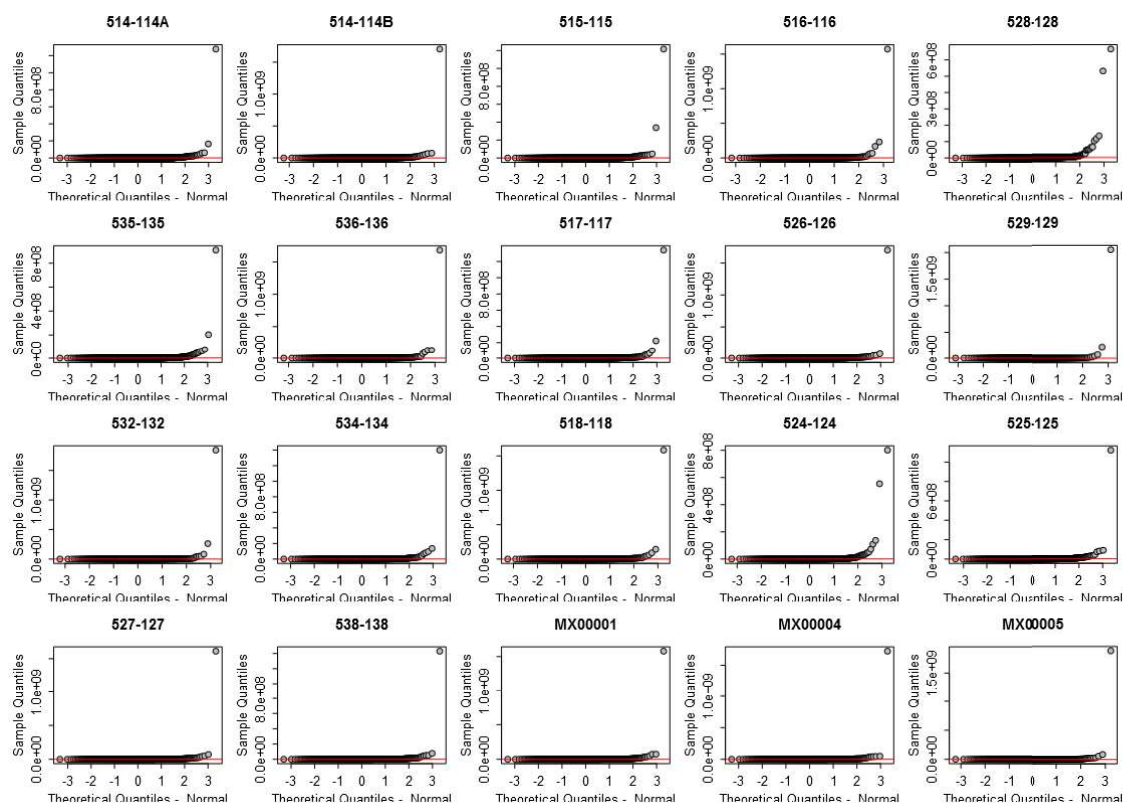


Figure 8.3.5: Q-Q-Plots for exosomal proteome sample data. Sample data origins from exosomes extracted from sample urine and analyzed via LC-MS in DIA mode. Sample identifiers are given. A diagonal line in a Q-Q-plot represents normal distribution. Any other graphical representation requires “non-parametric” data handling.

PCA Bi-Plot (54.2%)

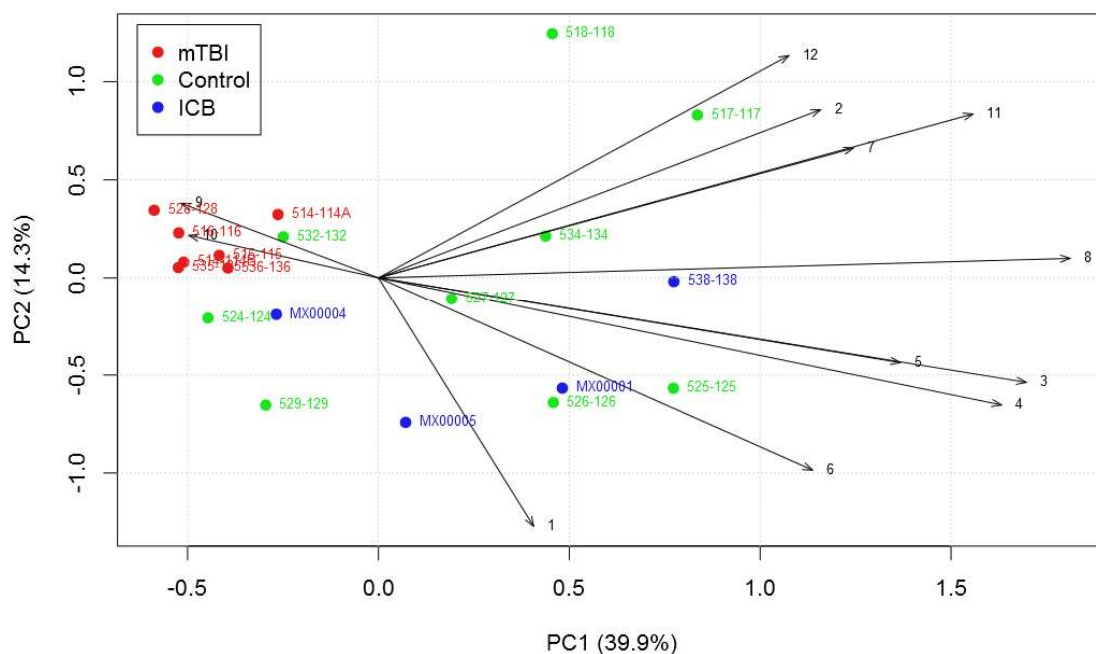


Figure 8.3.6: PCA Bi-Plot for filtered exosomal proteins. The principal components used for the plot (PC1 – 39.9% and PC2 – 14.3%) account for a total of 54.2% of the dataset variation. Data reflects group-wise separation between mTBI and controls as well as ICB samples. Arrows reflect protein-based added variance to the dataset and the length of the arrow the corresponding effect strength.

8.3.2 Extended Sample Cohort for Biomarker Qualification

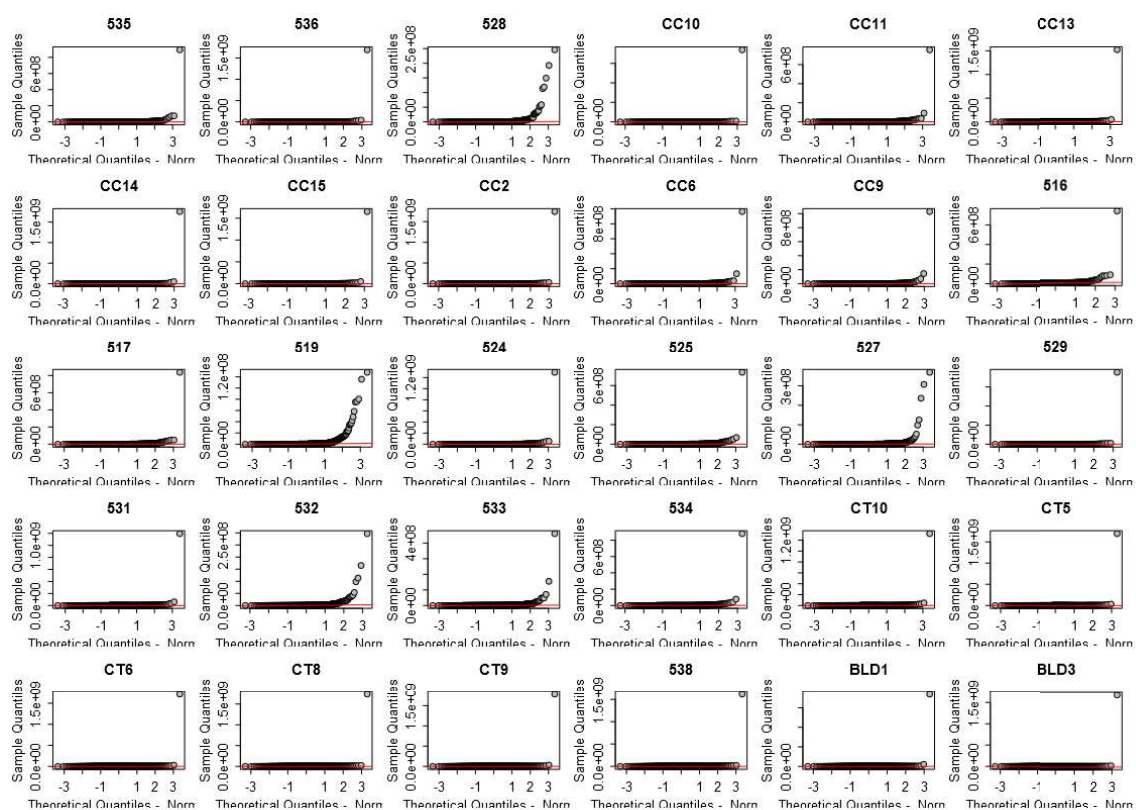


Figure 8.3.7: Q-Q-Plots of samples from the extended qualification sample cohort. Sample quantiles are plotted against theoretically calculated quantiles that represent normal distribution. Once, a diagonal line is visible (correlation between theoretical and real quantiles) normal distribution rules apply. Otherwise (non-diagonal line), non-parametric statistical principles need to be used.

8.3.3 Disease and Function Analysis

Table 8.3.1: List of associated “Diseases and Functions”. List retrieved from Ingenuity® IPA® software. Given are the names of the corresponding process, the associated p-value of relevance and the molecules (from the input list) which were associated with the disease or function.

Annotation	p-value	Molecules
abnormal morphology of cerebellum	2.01E-04	AXL,CNTN1,CP,GLB1,MYH10
abnormal morphology of fourth cerebral ventricle	6.47E-04	CP,MYH10
abnormal morphology of brainstem	6.67E-04	CP,GLB1,MYH10
abnormal morphology of Purkinje's layer of cerebellum	1.33E-03	CNTN1,CP,MYH10
Movement Disorders	1.45E-03	ACP1,APOD,CNTN1,CP,GLB1,HSPA1A/HSPA1B,HSPA8,PPP1CB
disorder of basal ganglia	1.76E-03	ACP1,APOD,HSPA8,KNG1,PPP1CB
edema of brain	2.19E-03	KNG1,PROCR
GM1 gangliosidosis	3.49E-03	GLB1
abnormal morphology of Golgi interneurons	3.49E-03	CNTN1
abnormal morphology of brain	3.92E-03	AXL,B4GALT1,CNTN1,CP,CSPG4,GLB1,MYH10
Huntington's Disease	6.12E-03	ACP1,APOD,HSPA8,PPP1CB
abnormal morphology of Purkinje cells	1.17E-02	CNTN1,CP
abnormal morphology of granule cell layer	1.22E-02	CNTN1,CP
neuritic dystrophy	1.39E-02	CLU
gait disturbance	1.61E-02	CNTN1,CP,GLB1
abnormal morphology of dendrites	1.64E-02	CNTN1,MYH10
abnormal morphology of neurites	2.06E-02	CNTN1,MYH10,PTPN13
damage of blood-brain barrier	2.08E-02	KNG1
abnormal morphology of nervous system	2.44E-02	AXL,B4GALT1,CNTN1,CP,CSPG4,GLB1,MYH10,PTPN13
neurodegeneration of hippocampal CA3 region	2.76E-02	GLB1
abnormal morphology of cerebral aqueduct	3.10E-02	MYH10
abnormal morphology of dilated third cerebral ventricle	4.78E-02	MYH10
abnormal morphology of retinal ganglion cells	4.78E-02	PTPN13
infarction of brain	4.78E-02	PROCR
demyelination	4.95E-02	CP,GLB1

8.4 Terminomal Biomarker Discovery

Table 8.4.1: Terminomal peptides of interest for diagnosing mTBI. Sequence is the observed semi-specific peptide sequence. The p-Value originates from the initial abundance comparison between mTBI and Controls jointly with ICB samples. Accession is the corresponding UniProt accession number, the peptide is associated with.

Sequence	p-Value	Assoc. Accession
FLQHKDDNPNLPR	0.00419376	P02768
VKVVVNFAPTIQE	0.00551578	Q7Z3B1
TCEPIQSVFFFSGDKYYR	0.01236189	P04004
NPPVQENFDVNK	0.01306114	P05090
DSENVQCQERDADPDTFFAK	0.01496447	P43652
YLNWYQQKPGK	0.01665896	P04431
LQGTLLGYR	0.01710804	P30530
EYCGVPGDGDDELLRFS	0.01710804	P02760
DYSVLLLLR	0.02221585	P02768
ADICTLSEKER	0.02398678	P02768
GKDTFYSLGSSLDITFR	0.02587886	O00187
VGHEALPLAFTQK	0.02587886	P01876
LVQPGGSLR	0.02721078	P01764
SVFPLAPSSK	0.02860144	P01857
SHGQDYLVGNK	0.03479291	Q16772
ASSLQSGVPSR	0.03479291	P01601
TQSPSSLSASVGDR	0.03650868	P01601
ASSYLSLTPEQWK	0.03921724	P0CG05
IPAVPPPTDLR	0.04111622	P02751
SALEVDETYVPK	0.04209456	P02768
WSGHCCLWDASVQVK	0.04209456	P07911
SLDGGFVYIAGK	0.04411042	P02787

Table 8.4.2: Terminomal peptides of interest for diagnosing ICB. Sequence is the observed semi-specific peptide sequence. The p-Value originates from the initial abundance comparison between ICB and Controls jointly with mTBI samples. Accession is the corresponding UniProt accession number, the peptide is associated with.

Sequence	p-Value	Assoc. Accession
KETCFAEEGKK	0.00014143	P02768
FGDEEERR	0.00060147	Q9JLV1
DSENVQER	0.00100033	P43652
TLGDNTAANNVR	0.00551667	Q7Z5L0
SPSGCGGTLYGDR	0.00902067	O60494
KIYPTVNCQPLGMIS	0.00965769	P01042
KESNEELTESCETK	0.01535423	P01042
YTSQEDLVEKK	0.01858982	P05156
PAQGNPNEEVAR	0.01858982	P15328
TGECTATVGKR	0.02106503	P01042
GSEAINAPGDN	0.02240697	Q9NZP8
AVLHVHGGG	0.02382262	P98160
KGGLIAYR	0.02531521	P49441
LQGSMLKPSSL	0.02688806	P01133
LAQELPQQLTSPGYPEPYGK	0.02770565	Q9NZP8
HQTVPQNTGGKNPD	0.02854458	P02787
GDEELLRFSN	0.02854458	P02760
TNPEDIYPSN	0.03028826	P16070
LWPWPQNFQTSQQR	0.03307506	P06865
FQLFGSPSGQKDLLF	0.04524561	P02788

8.5 Discovery of Biomarkers for NDs in CSF

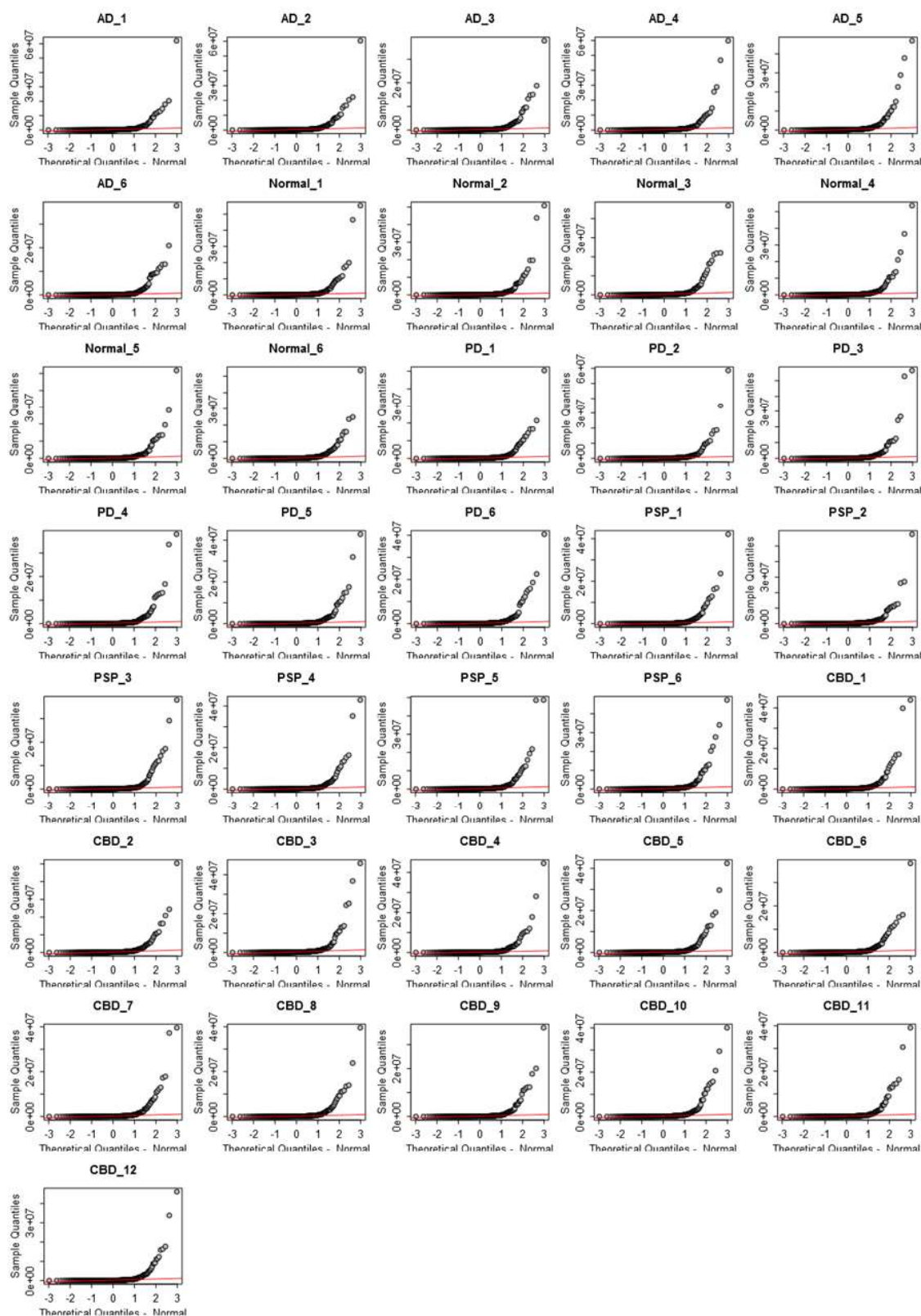


Figure 8.5.1: Q-Q-Plots for CSF samples analyzed by DIA. In this analysis, the theoretical quantiles (based on normal distribution) are plotted against the sample specific quantiles. With a strong correlation between the theoretical and the sample quantiles (diagonal line), an underlying normal distribution is evident. Here, none of the plots features a correlation that proves normal distribution. Non-parametric conditions apply.

8.5.1 PSP-centered Biomarker Discovery

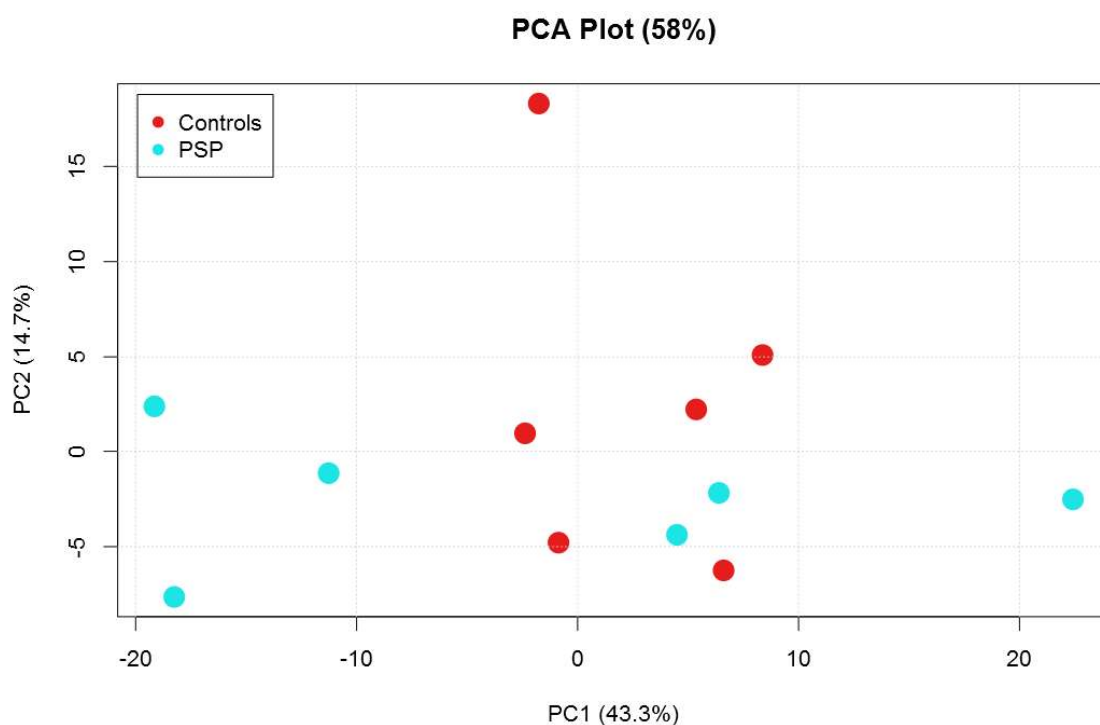


Figure 8.5.2: PCA plot of all proteins for PSP and Control samples. PCA plot is based on all quantified protein entries. Here, the combination of PC1 (43.3 %) and PC2 (14.7 %) account for a total of 58 % of the dataset variation. A first group-based separation (Controls [red] vs PSP [blue]) can be visually deviated.

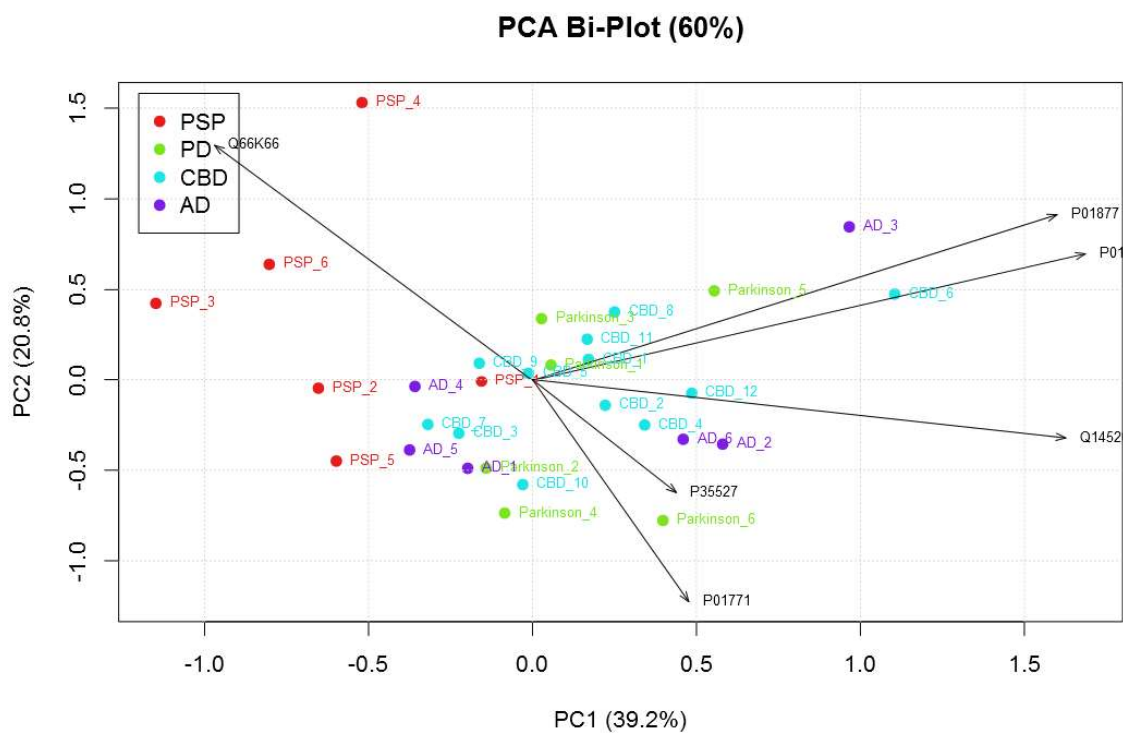


Figure 8.5.3: PCA Bi-Plot for PSP sample stratification. Basis for the PCA plot calculation are significantly differently abundant proteins, discovered by comparing PSP samples to the other neurodegenerative CSF samples. First two PCs account for 60 % of the dataset variance. Protein Q66K66 singly causes the spread of PSP samples away from the others.

8.5.2 PD-centered Biomarker Discovery

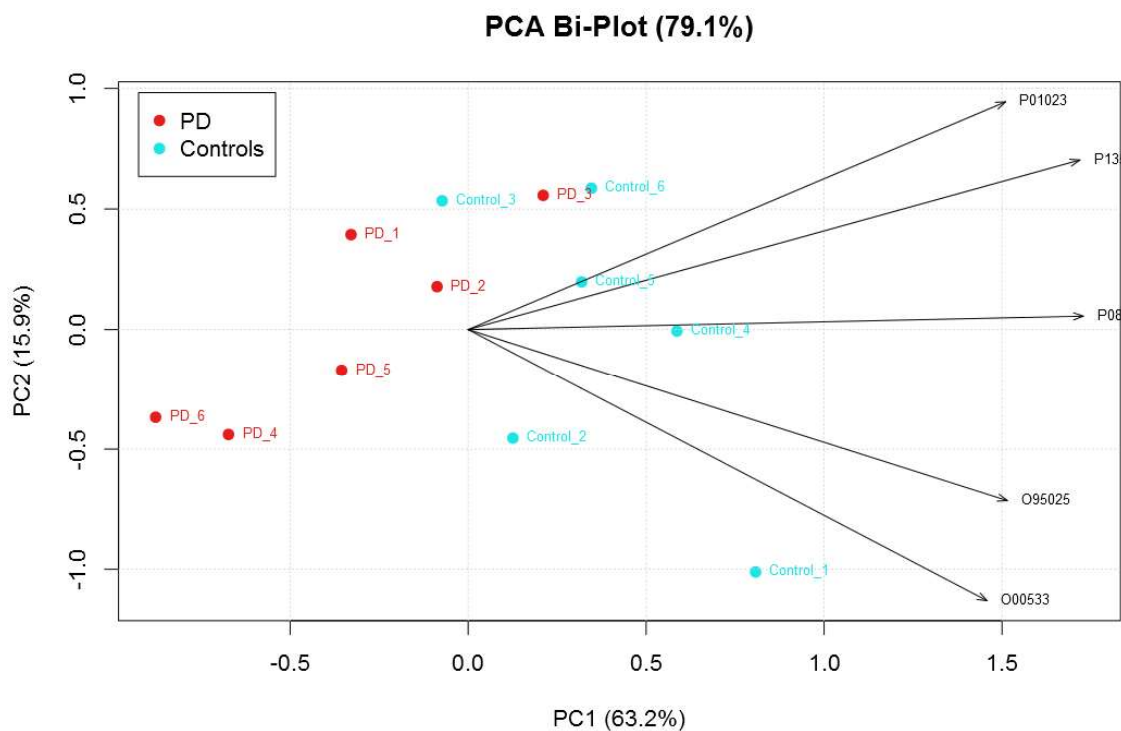


Figure 8.5.4: PCA Bi-plot for PD biomarker discovery. Basis of this PCA analysis demonstrate observed significantly different abundant proteins ($p < 0.05$). PCA loadings are shown via black arrows. PCA plot accounts for 79.1 % of the entered data variance, while PC 1 already accounts for 63.2 %. Segregation pattern can be segregated.

8.5.3 AD-centered Biomarker Discovery

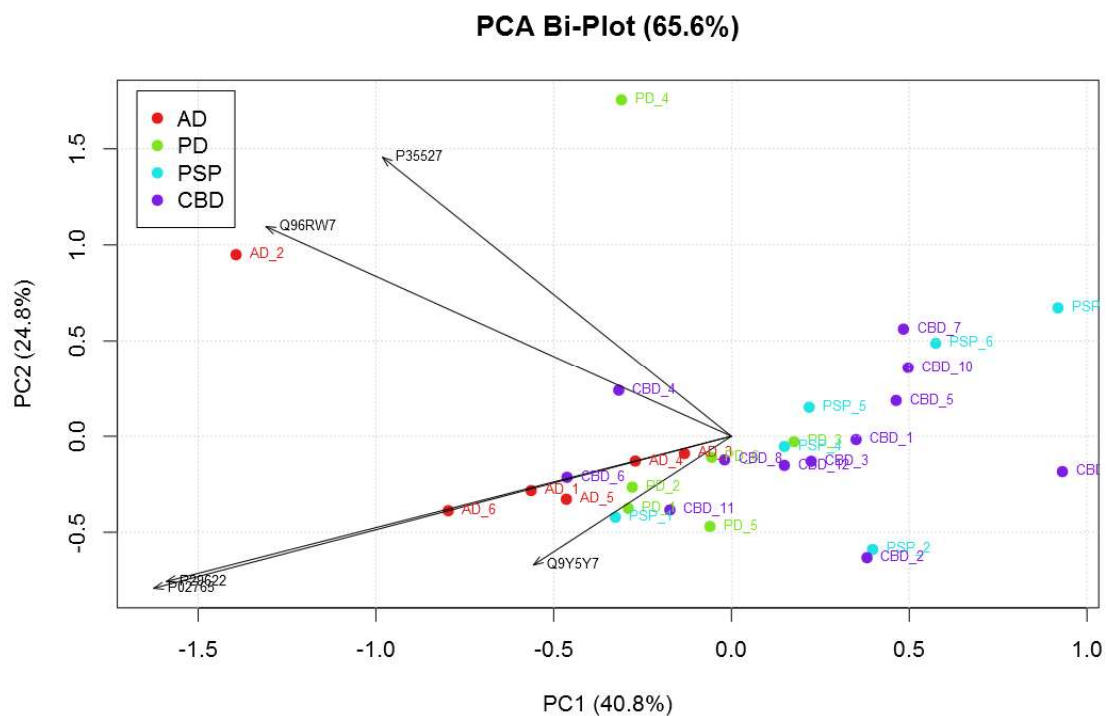


Figure 8.5.5: PCA Bi-plot for AD stratification. Arrows indicate PCA loadings. The plot is based on the seven identified significantly different proteins. AD samples start segregating along the vectors of three proteins.

8.5.4 CBD-centered Biomarker Discovery

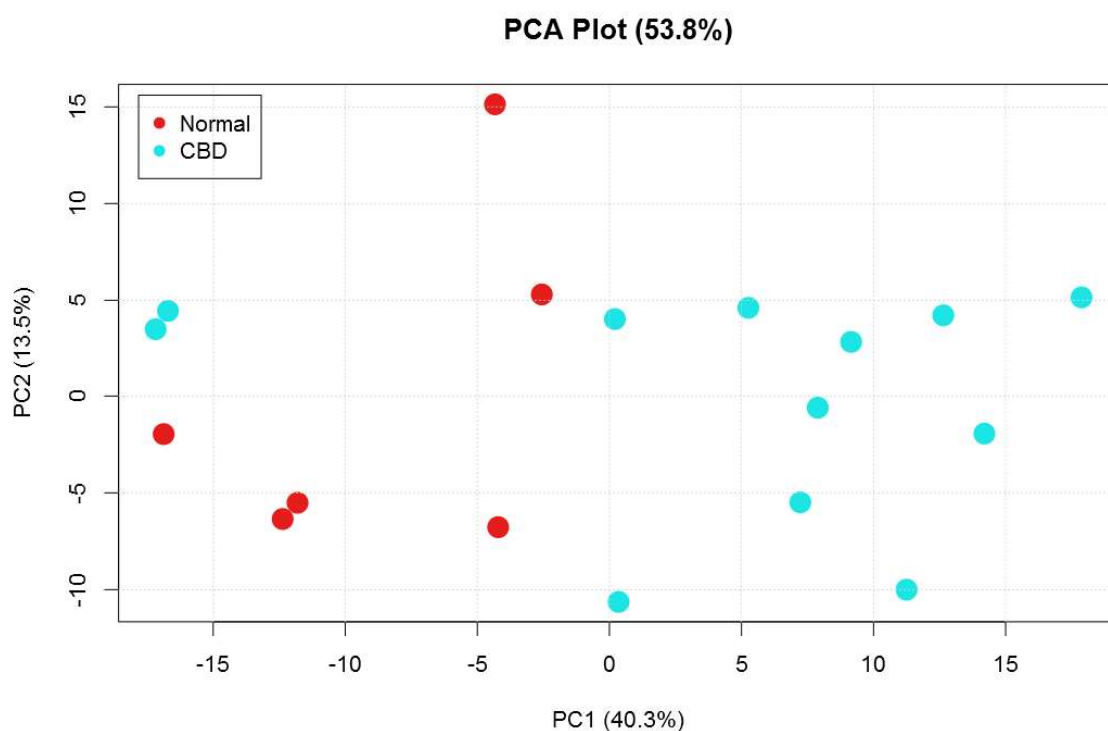


Figure 8.5.6: Unsupervised PCA Plot for CBD biomarker discovery. The basis for this PCA plot is the whole protein list. This list features 118 proteins that show CBD based significant protein abundance changes. The first principal component PC 1 accounts for 40.3 % of the whole dataset. All but two CBD samples segregate and form a group-based cluster.

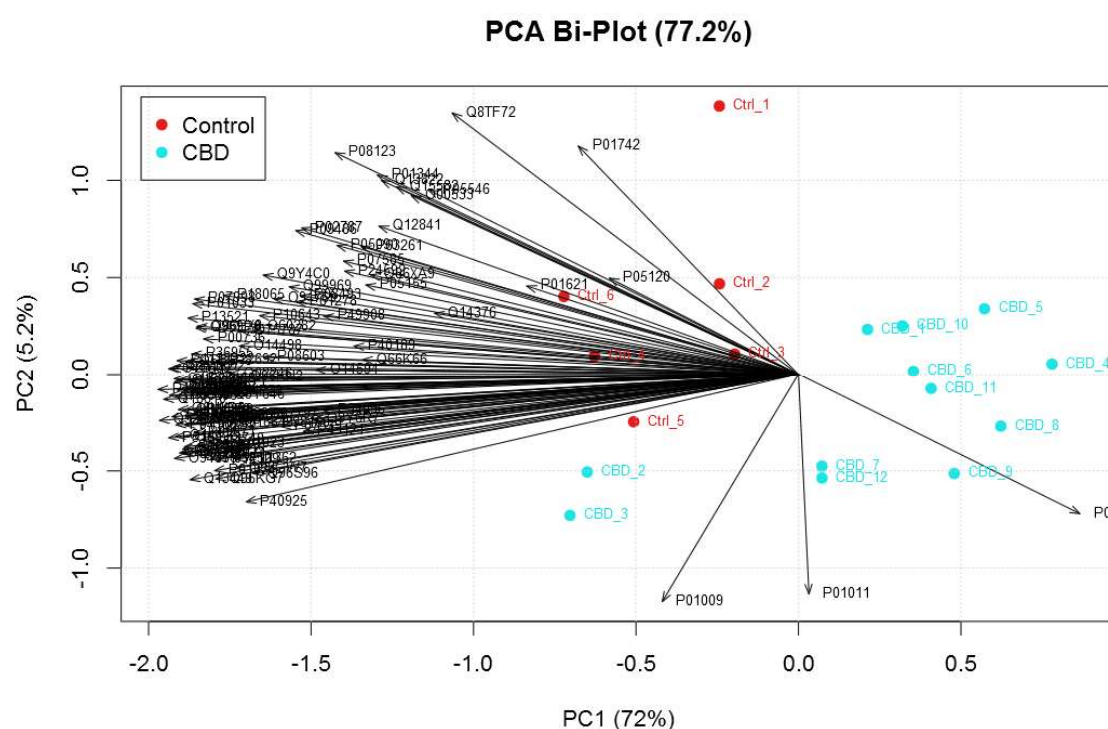


Figure 8.5.7: Supervised PCA Bi-plot for CBD biomarker discovery. The basis for this plot are 118 proteins observed with a significantly different protein abundance between control samples and CBD samples. Arrows indicate the PCA loadings and clearly show three proteins that cause the CBD related segregation.

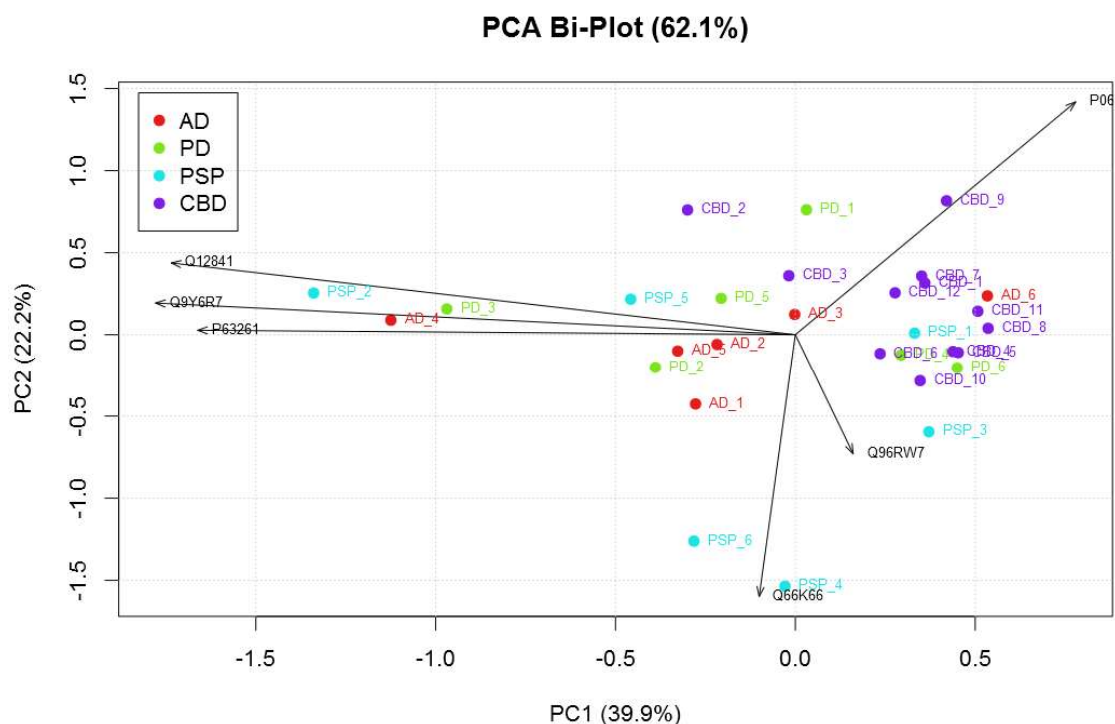


Figure 8.5.8: Stratification PCA Bi-Plot CBD vs NDs (1). This plot is based on PC1 (39.9 %; x – axis) and PC2 (22.2 %; y – axis). BiPlot accounts for a total of 62.1 % of the dataset variance. For stratification the control samples were not considered in this analysis. A certain CBD-based segregation tendency can be seen, nonetheless a clear segregation is not present.

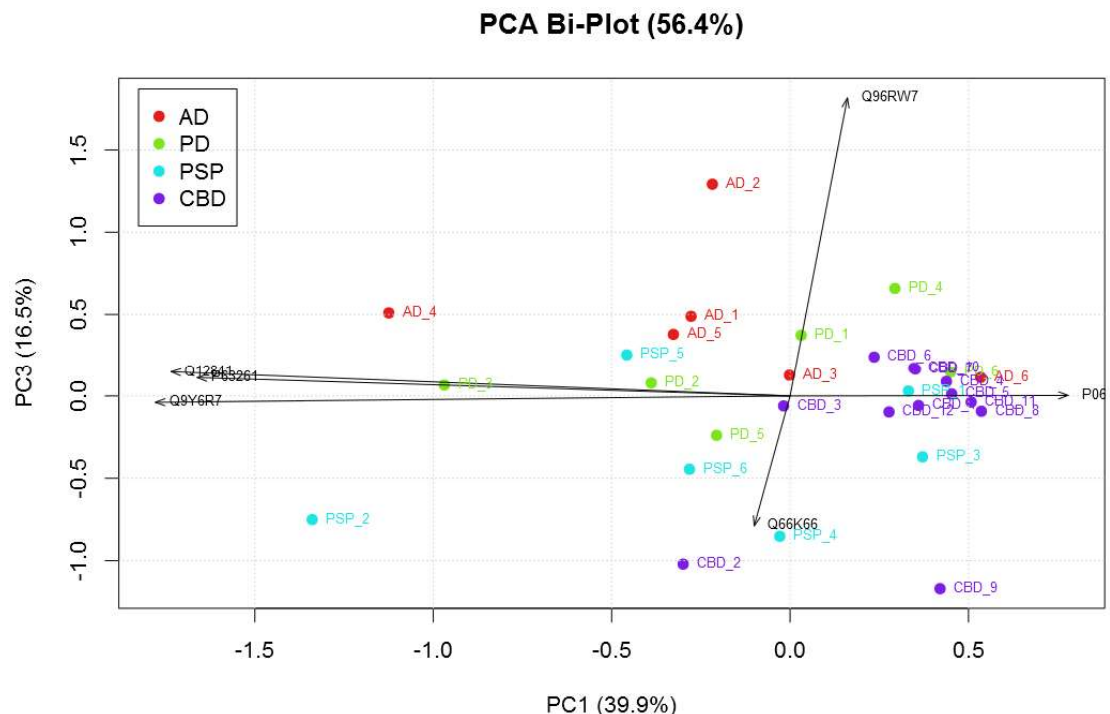


Figure 8.5.9: Stratification PCA Bi-Plot CBD vs NDs (2). The plot is based on the PC1 (39.9 %; x – axis) and PC3 (16.5 %; y –axis). A total of 56.4 % of the dataset variance is plotted via the present principal component combination. Control samples are disregarded in this type of analysis. Segregation pattern of CBD samples can be deviated, however does not demonstrate clear character. Major segregation strength along two vectors.

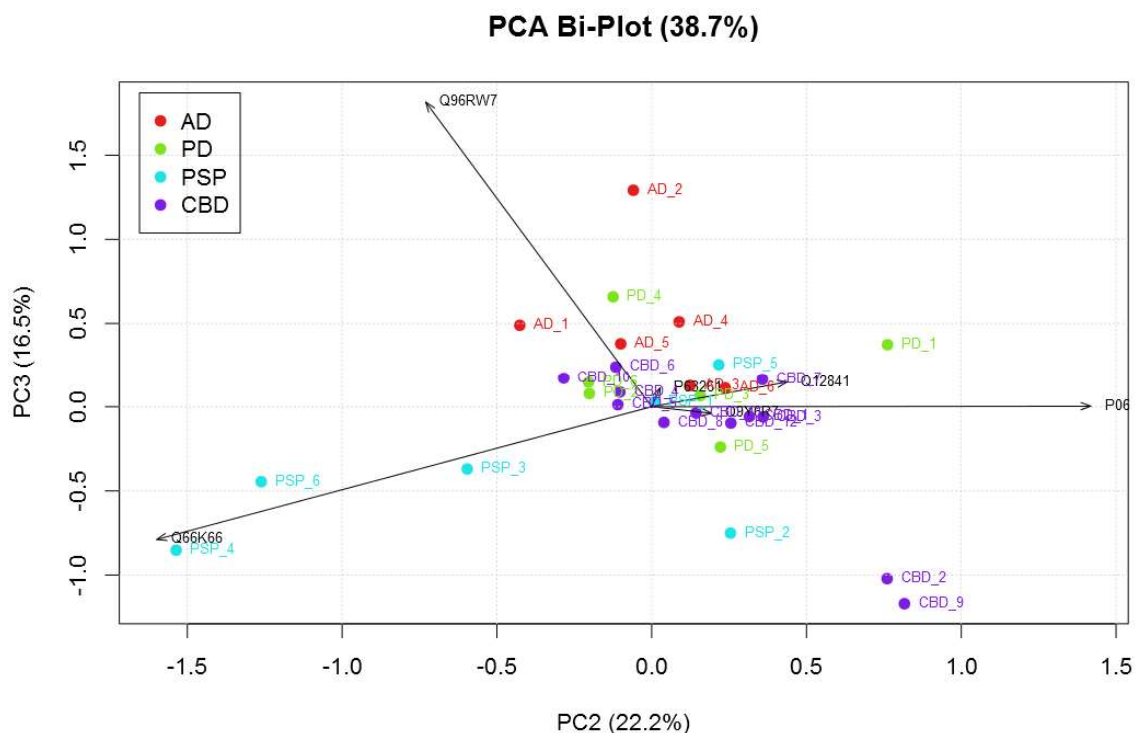


Figure 8.5.10: Stratification Bi-Plot CBD vs ND (3). The plot is based on the combination of PC2 (22.2 %; x – axis) and PC3 (16.5 %; y – axis). In total, a dataset variance of 38.7 % is plotted. A spread-out of two CBD samples is observable. Otherwise, rather PSP samples spread along one vector. CBD samples cluster together close to the origin. No specific segregation pattern can be deviated.

9. Curriculum Vitae of the Author

Sebastian T. Berger



Franz-Lehar-Str. 1

D-76275 Ettlingen

+49 (0)151 . 555 . 750 .97

Sebastian_Berger@gmx.de

04.04.1986

PhD Thesis

05 / 2013 – 01 / 2017 **Boston Children's Hospital and Harvard Medical School, USA**
Scope of research:
Clinical Proteomics, Mass Spectrometry, Urine Analysis.
Department of Pathology, Proteomics Center at Boston Children's Hospital, Director: Prof. Dr. Hanno Steen.
PhD supervisor: Prof. Dr. Bernhard Küster, Technical University of Munich; Chair of Proteomics and Bioanalytics.

Education

10 / 2009 – 11 / 2011 **Technical University Munich, Germany**
Study of Molecular Biotechnology
Master Thesis: "Cancer biomarker discovery by chemical and quantitative proteomics"
Chair of Proteomics and Bioanalytics, Chairman: Prof. Dr. Bernhard Küster

10 / 2006 – 03 / 2010 **Technical University Munich, Germany**
Study of Molecular Biotechnology
Bachelor Thesis: "Untersuchung zur Regulation insektizider Toxigene in Yersinia enterocolitica"
Chair of Microbial Ecology, Chairman: Prof. Dr. Siegfried Scherer

09 / 2002 – 07 / 2005 **Bertha-von-Suttner School, Ettlingen, Germany**
Abitur
Biotechnical Secondary School

Professional Experience

02 / 2017 – today **Agilent Technologies, Waldbronn, Germany**
Application Test Engineer
Responsibility: Preparative workflow-based (LC-MS) application testing and application defect management.

01 / 2012 – 04 / 2013 **Boston Children's Hospital and Harvard Medical School, USA**
Internship in Peptide Fractionation
Department of Pathology; Proteomics Center at Boston Children's Hospital, Director: Prof. Dr. Hanno Steen

08 / 2012 **AB Sciex, Framingham, Massachusetts, USA**
Administrator training for maintaining a TripleToF 5600

08 / 2007 – 10 / 2007 **Agilent Technologies, Waldbronn, Germany**
Summer Internship
Responsibility: assembly of, as well as troubleshooting of gas- and liquid- chromatographs (GC / LC)

 Patent

01 / 2016 *Patent application:* MStern blotting - An improved method of sample preparation for mass spectrometry-based proteomics. Pub. No. **WO/2016/112253**; Int. Appl. No. *PCT/US2016/012591*

 Publications

- **Berger, S.T.**, Ahmed, S., Muntel, J., Cuevas Polo, N., Bachur, R., Kentsis, A., Steen, J. and Steen, H. (2015). MStern blotting - high throughput PVDF membrane-based proteomic sample preparation for 96-well plates. *Molecular & Cellular Proteomics*. DOI: 10.1074/mcp.O115.049650
- Muntel, J., Xuan, Y., **Berger, S.T.**, Reiter, L., Bachur, R., Kentsis, A., Steen, H. (2015). Advancing Urinary Protein Biomarker Discovery by Data-independent Acquisition on a Quadrupole-Orbitrap Mass Spectrometer. DOI: 10.1021/acs.jproteome.5b00826
- Oh, D., Dowling, D.J., Ahmed S., Choi, H., Brightman, S., Bergelson, I., **Berger, S.T.**, Sauld, J. F., Pettengill, M., Kho1, A.T. Pollack, H.J., Steen, H. and Levy, O. (2016). Adjuvant-induced human monocyte secretome profiles reveal adjuvant- and age-specific protein signatures. DOI: 10.1074/mcp.M115.055541
- Davis-Anderson, K.L., **Berger, S.T.**, Lunde, E.R., Naik, V.D., Johnson, G.A., Steen, H., Ramadoss, J. (2017). Placental Proteomics Reveal Insights into Fetal Alcohol Spectrum Disorders. DOI: 10.1111/acer.13448

 Talks & Presentation

06 / 2015 *Conference of the American Society of Mass Spectrometry (ASMS)*; Poster presentation: MStern Blot – High Throughput PVDF Membrane-Based Proteomic Sample Preparation for 96-well Plates.

03 / 2017 *US HUPO*; Conference talk: Application and Optimization of “MStern Blot” for Translational Proteomics.

03 / 2017 *US HUPO*; Lightning talk: Getting a grip on what determines the composition of urinary proteomes.

03 / 2017 *US HUPO*; Poster presentation: Getting a grip on what determines the composition of urinary proteomes.

 Research Experience

Proteomic workflows

- Execution of proteomic protocols
- Optimizing of peptide separation and fractionation
- MS-based protein and peptide quantification
- Protein characterization
- Mass Spectrometry

- | | |
|--------------------------|--|
| Microbial workflows | <ul style="list-style-type: none">- Competent handling of microbes (S1, S2)- Targeted gen manipulation- Over expression of proteins in bacteria- Analysis of gen expression (e.g. promoter assay) |
| Bioanalytical techniques | <ul style="list-style-type: none">- HPLC-Analysis (micro- and nano-flow)- Immunoassay-Techniques (e.g. ELISA, Western Blot and IP)- Supervision of coworkers to self dependently trouble-shoot and diagnosing issues on HPLC and LC-MS systems |
| Mass spectrometers | <ul style="list-style-type: none">- AB Sciex TripleToF 5600+- Thermo Fisher Q Exactive |

Additional Competences

- | | |
|--------------------|---|
| Computer expertise | <ul style="list-style-type: none">- MS Office (Word, Excel, PowerPoint, Outlook, etc.)- Proteomic search algorithms (ProteinPilot, Mascot, etc.)- Programming (PHP, R, HTML, MySQL, MS-SQL)- Statistical data analysis (IBM SPSS, R, etc.)- Server setup and administration- Setup and maintenance of standalone NAS systems |
|--------------------|---|

Languages

- German: Native speaker
- English: Full professional proficiency
- French: Advanced proficiency
- Dutch: Basic proficiency

Interests and leisure activities

Travelling: Get to know new cultures

Sports: Dragon boat; Judo; Scuba diving

01 / 2014 – 12 / 2014 Board member of the „Boston 1 Dragon Boat Team“ in Boston, MA, USA.

10 / 2006 – 10 / 2009 Board-Administrator during the Bachelor studies in Munich

05 / 2016 – today Website administrator for Fontanherzen e.V.

Ettlingen, August 2017



Sebastian T: Berger