



Blood based biomarkers for the identification of Alzheimer's disease using proteomics approaches

Author:

Virk, Gurjeet Kaur

Publication Date:

2022

DOI:

<https://doi.org/10.26190/unsworks/24110>

License:

<https://creativecommons.org/licenses/by/4.0/>

Link to license to see what you are allowed to do with this resource.

Downloaded from <http://hdl.handle.net/1959.4/100403> in <https://unsworks.unsw.edu.au> on 2024-05-04

Blood based biomarkers for the identification of Alzheimer's disease using proteomics approaches

Gurjeet Kaur Virk



A thesis submitted for the degree of Doctor of Philosophy

School of Psychiatry

Faculty of Medicine

University of New South Wales

January 2022

Supervisors:

Professor Perminder Sachdev

Dr Anne Poljak

THE UNIVERSITY OF NEW SOUTH WALES SYDNEY

Surname/Family Name : **Virk**

First Name : **Gurjeet Kaur**

Abbreviation for degree : **PhD**

Faculty : **Faculty of Medicine**

School : **School of Psychiatry**

Thesis Title : **Blood based biomarkers for the identification of Alzheimer's disease using proteomics approaches**

Alzheimer's disease (AD) is a multifactorial neurodegenerative disease characterised by dysregulation of various cellular and molecular processes. Apart from environmental and lifestyle effects, genetic variations of the apolipoprotein E (*APOE*) gene plays a significant role in AD risk and progression, but these factors are poorly understood from a mechanistic perspective. In chapter 2, a meta-analysis of blood and CSF biomarkers of AD was performed, noting that the range of biomarkers studied has been restricted to a handful of classical proteins ($A\beta$ and tau) and heavily focused on CSF. Still, more research is needed to establish robust blood tests to complement CSF or imaging tests for non-invasive testing options. Mass spectrometry significantly outperforms conventional antibody-based approaches such as ELISA and western blotting in specificity and quantification of low abundant proteins. Plasma proteomics has historically been limited by the lack of throughput and sensitivity, owing mainly to the complexity of the plasma samples. In chapter 3, I have developed a method for determining the fractionation strategy that provides in-depth plasma proteome coverage identifying 4,385 total proteins. This work demonstrates that simpler and faster approaches can provide substantial proteome coverage in conventional biochemistry laboratories. In chapter 4, I performed label-free proteomics analysis on plasma samples from clinical cohorts, using the newly developed fractionation method. Longitudinal and cross-sectional analyses of normal ageing and ageing with progression to MCI and AD were performed, based on plasma proteomic changes in the Sydney Memory and Ageing Study cohort. A replication cohort was used in my chapter 5, the Australian Imaging, Biomarkers and Lifestyle study, which also included *APOE* ϵ 3 and ϵ 4 allele carriers. This additional information facilitated plasma proteome profiling to understand the impact of *APOE* ϵ 3 and *APOE* ϵ 4 carriers on AD dementia. Apart from comparing the effect of *APOE* genotypes on the AD proteome, I have confirmed a panel of reliable AD biomarkers that are consistently changing in both cohorts. In conclusion, I have successfully developed and applied MS-based fractionation methods for in-depth plasma proteome coverage of age, cognition and disease-related changes. Finally, a list of 44 plasma biomarkers consistently dysregulated in both AD cohorts presents a promising foundation for future clinical studies.

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 words abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).

Signature

Date: 13/01/22

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years must be made in writing. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

Copyright Statement

'I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 370 words abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only). I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted, I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed:

Date: 13/01/2022

Authenticity Statement

'I certify that the library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed:

Date: 13/01/2022

Thesis submission for the degree of Doctor of Philosophy

Thesis Title and Abstract

Declarations

Inclusion of Publications
Statement

Corrected Thesis and
Responses

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in the candidate's thesis in lieu of a Chapter provided:

- The candidate contributed **greater than 50%** of the content in the publication and are the "primary author", i.e. they were responsible primarily for the planning, execution and preparation of the work for publication.
- The candidate has obtained approval to include the publication in their thesis in lieu of a Chapter from their Supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis.

The candidate has declared that **some of the work described in their thesis has been published and has been documented in the relevant Chapters with acknowledgement.**

A short statement on where this work appears in the thesis and how this work is acknowledged within chapter/s:

Chapters 2 and 3 are the part of the articles which I have published previously as primary author. I was responsible primarily for the planning experiments, execution and preparation of the articles for publication. Acknowledgement of the work of other authors of these articles has been made at the beginning of each chapter as well as in the acknowledgement section of each article provided in the appendix section of the thesis.

Candidate's Declaration



I declare that I have complied with the Thesis Examination Procedure.

Abstract

Alzheimer's disease (AD) is a multifactorial neurodegenerative disease characterised by various cellular and molecular processes dysregulation. Apart from environmental and lifestyle factors, genetic polymorphisms of the apolipoprotein E (*APOE*) gene and other risk genes play significant roles in AD pathogenesis. Still, these factors are poorly understood from a mechanistic perspective. In chapter 2, a meta-analysis of blood and cerebrospinal fluid (CSF) biomarkers of AD was performed, noting that the range of biomarkers studied has been restricted to a handful of classical proteins (β -amyloid and tau) and heavily focused on CSF. Still, more research is needed to establish robust blood tests to complement CSF or imaging tests for non-invasive testing options. Mass spectrometry significantly outperforms conventional antibody-based approaches such as ELISA and Western blotting in specificity and quantification of low abundance proteins. Plasma proteomics has historically been limited by the lack of throughput and sensitivity, owing mainly to the complexity of the plasma samples. In chapter 3, I evaluated several fractionation strategies, providing an in-depth plasma proteome coverage identifying 4,385 total proteins. This work demonstrates that more straightforward and faster approaches can provide substantial proteome coverage in conventional biochemistry laboratories. In chapter 4, I performed label-free proteomics analysis on plasma samples from clinical cohorts, using the newly developed fractionation method. Longitudinal and cross-sectional analyses of normal ageing and ageing with progression to MCI and AD were performed based on plasma proteomic changes in the Sydney Memory and Ageing Study cohort. A replication cohort was used in chapter 5, the Australian Imaging, Biomarkers and Lifestyle study, including *APOE* ϵ 3 and ϵ 4 allele carriers. This additional information facilitated plasma proteome profiling to understand the impact of *APOE* ϵ 3 and *APOE* ϵ 4 carriage on AD dementia. Apart from comparing the effects of *APOE* genotypes on the AD proteome, I have confirmed a panel of reliable AD biomarkers that are consistently altered in both cohorts. In conclusion, I have successfully developed and applied MS-based fractionation methods for in-depth plasma proteome coverage of ageing, cognition and disease-related changes. Finally, a list of 44 plasma biomarkers consistently dysregulated in both AD cohorts presents a promising foundation for future biomarker studies.

Originality Statement

‘I hereby declare that this submission is my work, and to the best of my knowledge, it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Furthermore, any contribution made to the research by others with whom I have worked at UNSW or elsewhere is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or style, presentation and linguistic expression is acknowledged.’

Signed:

Date: 13/01/2022

Acknowledgements:

I would like to thank my primary supervisor Professor Perminder Sachdev who provided a high level of guidance, support and mentorship during my candidature.

I sincerely thank my co-supervisor, Dr Anne Poljak, for her expert advice and guidance in all aspects of my experimental work and support in editing my thesis and manuscripts. I am also highly grateful for her encouragement and personal support during the years of my PhD. In addition, I wish to thank A/Prof Mark Raftery for providing me with desk and lab space during my candidature to complete my mass spectrometry work. Finally, I would also like to extend many thanks to all other members of the Bioanalytical Mass Spectrometry Facility at the Mark Wainwright Analytical Centre, UNSW, for their help in many aspects of mass spectrometry instrumentation and sample preparation, as well as for their kind encouragement, friendship and support during my PhD.

I want to mention my best friend and constant support, Syed Azmal Ali. I am grateful for his friendship, encouragement and understanding throughout my PhD. I also acknowledge Isham Kaur, Sujith Kumar and Samitha Fernando for their kind support and conversations, which helped me get through difficult times. I am thankful to have Somansh, Vasu, Jaibeer, Lovedeep and Bharti. I feel fortunate to have such amazing friends in my life. I want to mention my incredible housemates Navya, Soumya, Talha, Zahid, Midhat, Gunjan and Tavishi. I am thankful for all the meals and love I have received during my thesis writing. My sincere thanks also go to my colleagues Mabi, Marina, Fatemeh, Yue and Matthew for their helpful conversations.

I have no words to express the support and love my brother Jitender Singh has provided throughout my life and in PhD using his expertise in python and R language skills. Finally, I would like to extend my deepest gratitude to my family who have been a great source of understanding, compassion, love and encouragement during the highs and lows of my PhD. My father, Kulwant Singh and my mother, Savinder Kaur, are the first and foremost reasons I have finished my doctorate journey.

Publications and Presentations:

First Author Peer-Reviewed Publications

Gurjeet Kaur, Anne Poljak, Nady Braidy, John D Crawford, Jessica Lo, Perminder S Sachdev. Fluid biomarkers and *APOE* status of early onset Alzheimer's disease variants: A systematic review and meta-analysis. *Journal of Alzheimer's Disease*, 2020, 75: 827-843.

Gurjeet Kaur, Anne Poljak, Syed Azmal Ali, Ling Zhong, Mark J Raftery, Perminder Sachdev. Extending the depth of human plasma proteome coverage using simple fractionation techniques. *Journal of Proteome Research*, 2021, 20(2), 1261-1279.

In preparation

Gurjeet Kaur, Anne Poljak, Perminder Sachdev. Deep proteome analysis of plasma reveals novel biomarkers of Alzheimer's disease: A longitudinal study. Manuscript

Gurjeet Kaur, Anne Poljak, Colin L Masters, Christopher Fowler Perminder Sachdev. Impact of *APOE* ϵ 3 and ϵ 4 genotypes on plasma proteome signatures in Alzheimer's diseases. Under Review

Other first author Peer Reviewed publications

Gurjeet Kaur, Syed Azmal Ali, Pradip Behare. Phosphoproteomics reveals the systemic adaptation mechanism of *Lactobacillus fermentum* to acidic conditions. Manuscript submitted, 2022

Gurjeet Kaur, Syed Azmal Ali, Pradip Behare. Proteotranscriptomics analysis of bile salt adaptation of vaginal isolated *Lactobacillus* probiotics. Manuscript submitted, 2022

Collaborative Peer-Reviewed Publications

Deepti Mittal, **Gurjeet Kaur**, Parul Singh, Karmveer Yadav, Syed Azmal Ali. Nanoparticle-based sustainable agriculture and food science: recent advances and future outlook. *Frontiers in Nanotechnology*, 2020, Dec 4;2:10.

Syed Azmal Ali, Deepti Mittal, **Gurjeet Kaur**. In-situ monitoring of xenobiotics using genetically engineered whole-cell-based microbial biosensors: recent advances and outlook. *World Journal of Microbiology and Biotechnology*, 2021, 37(5).1-24.

Book chapters

Gurjeet Kaur, Syed Azmal Ali. Decoding the Nano-bio effects on the cellular expressions in plants. Elsevier Book (Under Book Review), 2022

Publications and Presentations

Sonia Batan, Harpreet Kaur, Swasti Rawal, Deepti Mittal, Parul Singh, **Gurjeet Kaur**, Syed Azmal Ali. System biology research to advance understanding of canine cancer. Springer Book. (Under book review), 2022

Ben Geoffrey A S, Jitender Singh Virk, Deepti Mittal, Parul Singh, **Gurjeet Kaur**, Syed Azmal Ali. Data-driven and AI approach for system-wide prediction of the drugable proteome to drug discovery in farm animals. Springer Book. (Under book review), 2022

Harpreet Kaur, Sonia Bhathan, Swasti Rawal, Parul Singh, Deepti Mittal, **Gurjeet Kaur**, Syed Azmal Ali. System biology-based understanding of gut microbiome role in cattle production and health. Springer Book. (Under book review), 2022

Daraksha Iram, Manish Singh Sansi, Chandhni P.R, Sameena Zhanab, Sikha Rana, Priya Kalyan, Kapil Singh Narayan, Deepti Mittal, **Gurjeet Kaur**, Syed Azmal Ali. Effect of metal oxide nanoparticles on biochemical pathways in plants. Springer Book. (Under book review), 2022

Conference Presentations

Gurjeet Kaur, Anne Poljak, Nady Braidy, Perminder S Sachdev. P3-226: CSF and blood biomarkers of early-onset Alzheimer's disease: A systematic review and meta-analysis. Alzheimer's & Dementia, 2018

Gurjeet Kaur, Anne Poljak, Nady Braidy, Perminder S Sachdev. P2-219: Method development and proteomic analysis of depleted plasma for biomarker discovery in neurodegenerative diseases. Alzheimer's & Dementia, 2019

Prizes-Awards

Kwan Fung & Yuet Ying Fund Research Award for research into the diagnosis, prevention and cure of dementia and Alzheimer's disease, 2021

Awarded UNSW Medicine ARC PGC Student award in recognition of efforts to continue researching and contribute to the UNSW higher degree research community during the COVID-19 pandemic, 2021

Young Investigator Award at Proteomics for system integrated bio-omics, one health and food safety, NDRI, Haryana, India, 2019

Young Investigator Award at DIAD-AAIC conference from Washington State University St. Louise, USA, 2018

UNSW Scientia PhD Scholarship based on exceptional academic record, 2017-2021

I. Table of Contents

I. TITLE	I
II. THESIS DISSERTATION SHEET	II
III. ABSTRACT	VI
IV. ORIGINALITY STATEMENT.....	VII
V. ACKNOWLEDGEMENTS.....	VIII
VI. PUBLICATIONS, PRESENTATIONS, AND PRIZES	IX
VII. TABLE OF CONTENTS	XI
VIII. ABBREVIATIONS	XV
XI. LIST OF TABLES	XVII
X. LIST OF FIGURES	XVIII

1. INTRODUCTION	1
1.1. DEMENTIA AND ALZHEIMER'S DISEASE	3
1.2. PATHOLOGY OF AD	3
1.2.1. A β pathology	4
1.2.2. Tau pathology	5
1.2.3. Treatment strategies	6
1.3. BIOMARKER AND BIOFLUIDS	7
1.4. METHODS FOR IDENTIFICATION OF PLASMA PROTEINS	9
1.4.1. ELISA (enzyme-linked immunosorbent assay)	9
1.4.2. Western blot (WB)	10
1.4.3. Immunohistochemistry (IHC)	11
1.4.4. Evolution of mass spectrometry	11
1.4.4.1. <i>What is mass spectrometry?</i>	12
1.4.4.2. <i>Mass spectrometry-based proteomics</i>	12
1.4.4.3. <i>Proteomic workflow (MudPIT approach)</i>	13
1.4.4.4. <i>Quantitation methods of MS</i>	14
1.5. BIOMARKERS OF AD PATHOLOGY	19
1.5.1. CSF biomarkers of AD	20
1.5.2. Blood biomarkers of AD	22
1.5.3. Imaging Biomarkers of AD	22
1.5.4. Additional biomarkers of AD	23
1.6. PROTEOMICS-BASED EMERGING BIOMARKERS.....	24
1.6.1. Different samples used for analysis of AD proteome	26
1.7. MAJOR CHALLENGES IN BLOOD-BASED PROTEOMICS TECHNOLOGIES	27
1.8. DISCOVERING THE EARLY BIOMARKERS OF AD PATHOLOGY.....	29
1.8.1. ADAD and EOAD	29

Table of Contents

1.8.1. Risk Factors for LOAD	30
1.9. PATHWAYS DYSREGULATED IN AD.....	31
1.9.1. Metabolic dysfunction	31
1.9.2. Clearance of misfolded proteins	32
1.9.3. Inflammation and oxidative stress.....	32
1.9.4. Post-translation modifications (PTMs).....	32
1.9.5. Concluding Remarks	34
2.0. AIMS AND RATIONALE.....	34
2. FLUID BIOMARKERS AND APOE STATUS OF EARLY ONSET ALZHEIMER'S DISEASE VARIANTS: A SYSTEMATIC REVIEW AND META- ANALYSIS.....	37
2.1. INTRODUCTION	38
2.2. METHODS	39
2.2.1. Search strategy	39
2.2.2. Data extraction	40
2.2.3. Meta-analyses	42
2.2.4. Data availability statement	43
2.3. RESULTS	43
2.3.1. Protein/peptide biomarkers	43
2.3.1.1. <i>Early onset Alzheimer's Disease (EOAD)</i>	43
2.3.1.2. <i>Early onset sporadic Alzheimer's disease (EOsAD)</i>	47
2.3.1.3. <i>Autosomal dominant Alzheimer's disease (ADAD)</i>	51
2.3.1.4. <i>APOEϵ4 allele enrichment in EOAD groups</i>	56
2.4. DISCUSSION	56
2.5. LIMITATIONS.....	60
2.6. CONCLUSIONS	61
3. EXTENDING THE DEPTH OF HUMAN PLASMA PROTEOME COVERAGE USING SIMPLE FRACTIONATION TECHNIQUES.....	62
3.1. INTRODUCTION	63
3.2. EXPERIMENTAL PROCEDURES	64
3.2.1. Depletion of high abundant proteins.....	65
3.2.1.1. <i>Plasma preparation for HU6 column</i>	65
3.2.1.2. <i>Plasma preparation for Hu14 column</i>	65
3.2.1.3. <i>Enrichment of low abundant proteins using bead-based depletion</i>	66
3.2.1.4. <i>Electrophoretic fractionation of low abundance proteins using 1D-SDS PAGE</i>	67
3.2.1.5. <i>In-gel trypsin digestion</i>	67
3.2.1.6. <i>In-solution protein digestion</i>	68
3.2.1.7. <i>Fractionation of low abundance proteins using C18 column (ZORABX extended-C18)</i>	68
3.2.2. Mass Spectrometry	69
3.2.3. Computational Analysis	69

Table of Contents

3.2.4.	Density and Multi Scatter plots	71
3.2.5.	Hierarchical Clustering (HCA).....	71
3.2.6.	Volcano plot and Principal Component Analysis (PCA).....	71
3.2.7.	Human Protein Atlas Mapping	72
3.3.	RESULTS	72
3.3.1.	Comparative proteomic data provided by plasma depletion and separation workflow	72
3.3.2.	Technical performance across workflows	77
3.3.3.	Comparison of different pre-fractionation and separation techniques	79
3.3.4.	Principal Component Analysis (PCA).....	79
3.3.5.	Volcano plot analysis.....	83
3.3.6.	Human protein atlas based tissue annotation	84
3.4.	DISCUSSION	85
3.5.	CONCLUSION	92

4. DEEP PROTEOME ANALYSIS OF PLASMA REVEALS NOVEL BIOMARKERS OF MILD COGNITIVE IMPAIRMENT AND ALZHEIMER'S DISEASE: A LONGITUDINAL STUDY94

4.1.	INTRODUCTION.....	95
4.2.	MATERIALS AND METHODS	96
4.2.1.	Experimental procedures	96
4.2.2.	Depletion of high abundant proteins using Human 14 (HU14).....	97
4.2.3.	Fractionation of low abundance proteins using 1D-SDS PAGE	97
4.2.4.	Computational Analysis	98
4.2.5.	Bioinformatics Analyses.....	99
4.3.	RESULTS	100
4.3.1.	Overview of proteomics study populations	100
4.3.2.	DEPs identified in longitudinal analyses of ageing regardless of diagnosis.....	110
4.3.3.	DEPs that change longitudinally with progression to MCI and AD from normal cognition ..	118
4.3.4.	Common plasma proteome changes in longitudinal AD and MCI groups	127
4.3.5.	Cross sectional proteome changes in AD and MCI – potential clinical biomarkers.....	131
4.3.6.	Common proteome changes in cross sectional analysis of AD and MCI	140
4.3.7.	Plasma proteome changes in preclinical AD– potential early biomarkers.....	147
4.4.	DISCUSSION	158
4.4.1.	Plasma proteome changes in ageing using a longitudinal analysis.....	158
4.4.2.	Plasma proteome changes associated with progression over 6 years.....	157
4.4.3.	Proteomic changes in clinically diagnosed AD and MCI relative to their age matched cognitively normal controls (potential clinical diagnostic markers.....	161
4.4.4.	Potential early disease markers; preclinical changes in AD	163
4.4.5.	Mechanisms of AD and MCI.....	164
4.4.6.	Picking candidate biomarkers specific to incipient MCI and AD and preclinical AD	166
4.5.	CONCLUSION	166

5. IMPACT OF APOE E3 AND E4 GENOTYPES ON PLASMA PROTEOME SIGNATURES IN ALZHEIMER'S DISEASES 167

5.1. INTRODUCTION..... 168

5.2. MATERIALS AND METHODS 169

5.2.1. Cohort and samples 169

5.2.2. Depletion of high abundant proteins using Human 14 (HU14) 169

5.2.3. Fractionation of low abundance proteins using 1D-SDS PAGE 170

5.2.4. Computational Analysis 170

5.2.5. Bioinformatics Analyses..... 171

5.3. RESULTS 172

5.3.1. Sample characteristics by *APOEε3* and *APOEε4* genotype..... 172

5.3.2. Overall plasma proteome changes in AD vs controls, all *APOEε3* carriers..... 176

5.3.3. Proteome changes seen in *APOEε4* carriers 184

5.3.4. Changes common to both AD genotypes (AD risk factor independent of *APOEε4*)..... 193

5.3.5. Proteins linked to *APOEε4* genotype in controls..... 194

5.4. DISCUSSION 200

5.4.1. Plasma proteome signatures associated with both *APOEε3* and *APOEε4*..... 200

5.4.2. Mechanisms related to *APOEε3* and *APOEε4*..... 203

5.4.3. Mechanism related to control E4 vs control E3..... 205

5.5. CONCLUSION 206

6. GENERAL DISCUSSION AND FUTURE DIRECTIONS 207

6.1. DISCUSSION AND CONCLUSION 208

6.2. FUTURE STUDIES 218

7. REFERENCES 220

8. APPENDIX 242

List of Abbreviations

Aβ	Amyloid beta
AD	Alzheimer's disease
ADAD	Autosomal dominant Alzheimer's disease
aMCI	Amnesic mild cognitive impairment
APOE	Apolipoprotein E
APP	Amyloid precursor protein
BBB	Blood brain barrier
CNS	Central nervous system
CSF	Cerebrospinal fluid
CTRL	Healthy controls
CV	Coefficient of variance
DEPs	Differentially expressed proteins
ELISA	Enzyme-linked immunosorbent assay
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionization
EOAD	Early onset Alzheimer's disease
EOsAD	Early onset sporadic Alzheimer's disease
FA	Formic acid
FBS	Fetal bovine serum
FDR	False discovery rate
FWHM	Full width at half maximum
GFP	Green Fluorescent Protein
GO	Gene ontology
GOI	Gene of interest
GSEA	Gene set enrichment analysis
HCD	Higher-energy collision dissociation
hpH-RPLC	high pH reversed-phase liquid chromatography
IAA	Iodoacetamide
LC	Liquid Chromatography
LC-MS	Liquid chromatography coupled to mass spectrometry
LOAD	Late onset Alzheimer's disease
LFQ	Label-free quantification
MAP	Microtubule-associated proteins
MeOH	Methanol
min	Minutes
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MS¹	First or full scan of precursor Ions in tandem mass spectrometry
MS²	Second scan of fragment ions in tandem mass spectrometry
m/z	mass over charge

List of Abbreviations

MRM	Multiple reaction monitoring
NC	Normal control
NfL	Neurofilament light
NFT	Neurofibril tangles
NMAD	N-methyl-D-aspartate
NSAF	Normalized spectral abundance factor
PBS	Phosphate-buffered saline
PD2.4	Proteome Discoverer 2.4
PCR	Polymerase chain reaction
PTM	Post-translational modification
P-tau	Phosphorylated-tau
PSEN1	Presenilin 1
PSEN2	Presenilin 2
QE HF	Q-Exactive High Field Mass Spectrometer
qPCR	quantitative polymerase chain reaction
rpm	Revolutions per minute
RT	Room temperature
RT-PCR	Reverse transcription-quantitative polymerase chain reaction
s	Seconds
SD	Standard deviation
SDC	Sodium deoxycholate
SDS	Sodium dodecylsulfate
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SPS	Synchronous precursor selection
SRM	Single reaction monitoring
TCEP	Tris(2-carboxyethyl)phosphine
TEAB	Triethylammonium bicarbonate
TFA	Trifluoroacetic acid
T-tau	Total-tau
WB	Western blot
XICs	Extracted ion chromatograms

List of Tables

Table 1.1: Gene mutations and malformed proteins as causative agents of neurodegenerative disease 2

Table 1.2: The advantages and limitations of SWATH-MS in comparison with DDA and targeted proteomics 18

Table 1.3: Identification of proteins from different sample types in AD 26

Table 2.1: Meta-analysis of CSF and serum/plasma biomarkers of EOsAD and ADAD 44

Table 3.1: Number of proteins, peptides and spectra identified in each fractionation approach 74

Table 3.2: Mass-spectrometry-based plasma proteomic studies for comparison 86

Table 4.1: Details of the participant demographics which were included in our present study 101

Table 4.2: List of DEPs in all the longitudinal and cross sectional comparisons analysed 108

Table 4.3: The final list of longitudinal 71 DEPs in normal ageing 113

Table 4.4: The list of proteins uniquely differentially expressed in longitudinal AD 121

Table 4.5: The list of proteins uniquely differentially expressed in longitudinal MCI 124

Table 4.6: A total of 19 DEPs were common in both the longitudinal AD and MCI 129

Table 4.7: The list of proteins uniquely differentially expressed in cross-sectional AD 133

Table 4.8: The list of proteins uniquely differentially expressed in cross-sectional MCI 136

Table 4.9: The list of 18 DEPs were common in both the cross sectional AD and MCI 141

Table 4.10: DEPs common in both longitudinal and cross sectional AD comparisons 143

Table 4.11: Common DEPs in both longitudinal and cross sectional MCI comparisons 145

Table 4.12: The final list of 160 DEPs involved in preclinical AD 149

Table 4.13: A total of 15 DEPs commonly DEPs in age-matched incipient AD and preclinical AD 156

Table 5.1: Demographics of the participants. 173

Table 5.2: Final list of differentially expressed proteins (DEPs) in *APOEε4* and *APOEε3* 179

Table 5.3: List of unique 65 DEPs were identified in ADE3 relative to control E3 181

Table 5.4: This table shows the unique 65 DEPs were identified in ADE4 relative to control E4 188

Table 5.5: This table shows the unique 65 DEPs were identified in ADE4 relative to control E3 189

Table 5.6: A list of 28 DEPs were identified commonly in ADE3/CTRLE3 and ADE4/CTRLE3 190

Table 5.7: A list of 9 DEPs were identified commonly in ADE4/CTRLE4 and ADE4/CTRLE3 192

Table 5.8: A list of 34 unique DEPs were identified in control E4 relative to control E3 196

Table 6.1: Identification of 44 DEPs in AIBL and MAS cohorts replicated in both cohorts 215

List of Figures

Figure 1.1i: Pathophysiology of Alzheimer’s disease.	5
Figure 1.1ii: Protein aggregation mechanisms in AD are shown schematically.....	9
Figure 1.2: Workflows of discovery-based and targeted proteomics using mass spectrometry.	14
Figure 1.3: An overview of MS-based quantitative proteomics.	16
Figure 1.4: An overview of SILAC labelling protocol.....	17
Figure 1.5: An overview of TMT labelling	18
Figure 2.1: Flow chart depicting the selection process for the meta-analysis.....	40
Figure 2.2: Forest plots of early onset sporadic Alzheimer’s disease patients vs controls	48
Figure 2.3: Forest plots of early onset sporadic Alzheimer’s disease patients (tested negative for APP/PSEN mutation carriers) vs normal control individuals.	50
Figure 2.4: Forest plots of autosomal dominant Alzheimer’s disease patients vs normal control individuals	52
Figure 2.5: Funnel plots and Egger’s test	55
Figure 2.6: A systematic overview of the meta-analysis	56
Figure 3.1: Schematic representation: workflow used for method optimization of plasma fractionation and proteomics analysis.....	73
Figure 3.2: NuPAGE LDS gel profile of undepleted and depleted plasma	73
Figure 3.3: Venn diagram showing the protein overlap across all depletion methods	76
Figure 3.4: Density scatter plots of technical replicates for each method.	77
Figure 3.5: Multiscatter plot depicting the cross-correlation of all the methods	77
Figure 3.6: LFQ intensity of all the methods.....	78
Figure 3.7: Juxtaposed LFQ intensities for pair-wise comparison of C18 vs SDS PAGE fractionation.	79
Figure 3.8: Principal component analysis.....	81
Figure 3.9: Venn diagram showing top 100 protein dynamic range among depletion methods.....	81
Figure 3.10: Protein rank abundance plot based on LFQ intensities	82
Figure 3.11: Volcano plot Analysis	83
Figure 3.12: Proteins identified in our plasma proteome representative of the 11 main organ systems.....	84
Figure 4.1: NuPAGE LDS gel profile of and depleted plasma containing low abundant plasma proteins (LAP) from HU14 from all the individuals	98
Figure 4.2i: Scatter plot and regression analysis of abundance ratio of DEPs in both PD2.4 and scaffold ..	103
Figure 4.2ii: Density plot and regression analysis was plotted using normal control, MCI, and AD in both longitudinal and cross-sectional comparisons	104
Figure 4.2iii: Scatter plot was plotted between all 6 comparisons of normal control, MCI, and AD using PD2.4 and Scaffold.....	105
Figure 4.3: Proteome profiling and comparison of normal ageing, MCI, and AD in longitudinal and cross-sectional cohorts	106
Figure 4.4: Heatmap and volcano plot of ageing-related protein changes.....	111
Figure 4.5: DEPs from longitudinal comparisons were sorted into protein function classes	112
Figure 4.6: Heatmap and volcano plot using longitudinal analysis of MCI and AD-related protein changes	120

List of Figures

Figure 4.7: Common proteome changes in longitudinal and cross sectional MCI and AD	128
Figure 4.8: Heatmap and volcano plot using cross sectional analysis of MCI and AD-related protein changes	132
Figure 4.9: Heatmap and volcano plot using preclinical AD-related changes.....	148
Figure 4.10: Schematic model of the core Hippo pathway in mammals	159
Figure 5.1: Workflow of plasma proteome profiling and comparison of <i>APOEε3</i> and <i>APOEε4</i> genotype..	174
Figure 5.2i: Scatter plots and density plots comparing PD2.4 versus scaffold.....	175
Figure 5.2ii: Images of NuPAGE LDS gels of depleted plasma	176
Figure 5.3: Global analyses of proteomic changes in <i>APOEε3</i> and <i>APOEε4</i>	178
Figure 5.4: Unique protein changes in each comparison of <i>APOEε3</i> and <i>APOEε4</i>	179
Figure 5.4: Pie charts categorized the DEPs in each group into their biological processes and molecular pathways	186
Figure 5.5: Common protein changes in <i>APOEε3</i> and <i>APOEε4</i> and their functional categories	194
Figure 5.6: Gene ontology enrichments common to the ADE3/CTRLE3 and ADE4/CTRLE4 groups.....	199

Chapter 1

General Introduction

1. Neurodegenerative diseases - an introduction

Neurodegenerative disease is an umbrella term that describes a range of clinical conditions occur when degeneration of neurons in the brain or peripheral nervous system leads to loss of function over time and, in many cases, ultimately death¹. The type of disease or syndrome depends on the type of cell(s) and brain region(s) affected, the combinations of affected regions, and the types of pathological accumulations that result in significant heterogeneity of clinical presentations in various combinations. In addition, numerous molecular processes, including programmed cell death, inflammation, autophagosomal/lysosomal system, proteotoxic stress and its attendant abnormalities in the ubiquitin–proteasomal molecular “machinery”, and oxidative stress are shared by a variety of neurodegenerative diseases².

Neurodegenerative diseases are characterized by amyloidoses, tauopathies, transactivation response DNA binding protein 43 (TDP-43) proteinopathies, and α -synucleinopathies³. Accumulations of toxic proteins are caused by several factors, including specific gene mutations, some of which are familial. Familial variants often vary from the sporadic form of the disease, with an earlier age of onset being one feature of the familial forms⁴. Some of the genetic mutations associated with commonly occurring neurodegenerative diseases are presented in Table 1.1. These disorders, unfortunately, are little understood in terms of their pathophysiology. This poses several challenges, such as diagnostic inaccuracy in differentiating some neurodegenerative disorders, identifying changes in the often long presymptomatic phase, and developing potential biomarkers specific to the pre-clinical stages of the diseases.

Table 1.1 Gene mutations and malformed proteins as causative agents of neurodegenerative disease

Disease	Genes mutations causing Familial variant	Risk factor genes for late-onset variant	Proteins commonly associated with plaque or inclusion bodies
Alzheimer’s Disease (AD)	APP, PSEN1, PSEN2	<i>APOEϵ4</i> , <i>CLU</i>	APP fragments, excess tau phosphorylation
Vascular Dementia (VD)	<i>HTRA1</i>	<i>NOTCH</i>	microinfarcts, lacunar infarcts
Lewy Body Dementia	SNCA or SNCB	<i>APOE</i> , <i>SNCA</i> , and <i>GBA</i>	alpha-synuclein, aggregation, and Lewy body
Frontotemporal Dementia (FTD)	C9ORF72, MAPT and progranulin (GRN)	<i>TMEM106B</i>	Accumulation of Tau protein and TDP-43
Parkinson’s Disease (PD)	SNCA, PARK7, PINK1, or PRKN	<i>LRRK2</i>	α -synuclein forms aggregates called Lewy bodies
Amyotrophic lateral sclerosis (ALS)	SOD, C9orf72	<i>SOD1</i> , <i>TARDBP</i> , <i>FUS</i> , and <i>C9ORF72</i>	TDP-43, SOD1, and FUS

1.1 Dementia and Alzheimer's disease

Dementia describes a collection of symptoms associated with progressive cognitive impairment, often involving memory loss, caused by disorders affecting the brain. The ability to perform independent tasks and daily activities decline with disease progression, eventually diminishing independence and quality of life. The initial symptoms are often behavioural changes, forgetfulness, problems in executive function, language impairment, and visuospatial dysfunction. Dementia is usually caused by neurodegeneration, with the most common types of dementia being Alzheimer's disease (AD), vascular dementia, and Lewy body dementia⁵. AD accounts for up to 60-70% of all dementia cases, and according to the World Health Organization (WHO) 2019-2020 report, it is now estimated to be the third-leading cause of deaths worldwide after heart disease and cancer⁶. AD international reported that ~50 million people worldwide live with dementia (2020), of which ~60% live in low to middle-income countries⁷. In addition to the health burden, dementia is an immense economic burden on society. As the world's population ages, dementia emerges as a significant challenge internationally. The WHO has considered dementia a public health priority and implemented a global action plan aiming to increase dementia awareness, improve care, diagnosis, treatment, and support dementia research to accelerate the development of better diagnosis, prevention, and effective treatments.

1.2 Pathology of AD

AD is a progressive neurodegenerative disorder and the most prevalent cause of dementia. It was first described and named after a German psychiatrist and pathologist, Alois Alzheimer, in 1906. The most common sporadic AD form occurs in older people (>65 years), the late-onset Alzheimer's disease (LOAD), and accounts for 95% of all AD cases. In contrast, early-onset Alzheimer's disease (EOAD) accounts for about 5% of all AD cases⁸. The familial or autosomal dominant Alzheimer's disease (ADAD) represents about 1% of AD cases, and it is the more aggressive form due to specific gene mutations in one of the three genes: amyloid APP, PSEN1, and PSEN2⁹⁻¹². More than 350 mutations have been reported in these genes (<https://www.alzforum.org/mutations>). Sporadic AD also has strong genetic underpinnings, with several risk genes identified. Three alleles, i.e., *APOE* ϵ 2, *APOE* ϵ 3, and *APOE* ϵ 4, result in six possible genotypes (*APOE* 2/2, 2/3, 3/3, 2/4, 3/4, and 4/4). These three polymorphic alleles, i.e., ϵ 2, ϵ 3, and ϵ 4, have a worldwide frequency of 8.4%, 77.9%, and 13.7%,

respectively^{13,14}. Recent studies reported that approximately 65% of individuals with late-onset familial and sporadic AD bear the *APOE* ϵ 4 allele⁴. One copy of *APOE* ϵ 4 is associated with a threefold increase in disease risk, while two copies are associated with a more than tenfold increase in risk¹⁵.

1.2.1 A β pathology

One of the most commonly observed pathology-associated markers in AD is amyloid-beta (A β) peptide, whose complex lifecycle produces not only a plethora of chain length variants such as A β ₃₈, A β ₄₀, A β ₄₂, and A β ₄₃¹⁶, but also multimeric aggregate structures such as oligomers, fibrils, and plaque¹⁷. Of these, an increase in the production of the A β ₄₂ (42 amino acid long fragment), due to either mutation in the amyloid precursor protein (APP) or disruption in other molecular pathways (e.g., inflammation, oxidative stress, poor clearance due to associated vascular issues, and others), leads to aggregation of toxic A β ₄₂ variants, such as oligomers¹⁸. With disease progression, A β plaque eventually builds up in various brain regions. This process, known as the sink model, results in lower soluble A β ₄₂ as plaque/insoluble levels increase¹⁹. Additionally, A β has been shown to aggregate in two distinct reactions: non-metal-dependent association and metal-dependent association. Non-metallic A β aggregates produce soluble oligomers and amyloid fibrils, whereas metallic A β aggregates generate ionically bridged aggregates, covalently cross-linked oligomers, and precursors for non-metallic A β fibrillization²⁰. Deposition of A β initially results in the formation of A β oligomers and later fibrils and senile plaques. Our meta-analysis results showed a drop in CSF A β ₄₂ levels in EOAD variants, and this observation is consistent with previous meta-analytic data on LOAD CSF A β ₄₂²¹. According to current knowledge, A β oligomers are the most toxic variations, particularly synaptic function, whereas, A β plaque is likely a less harmful endpoint, in which A β aggregates with the cellular debris from dying neurons functionally much less problematic. Several cell mechanisms play an essential role in the degradation of misfolded proteins such as lysosomes, endosomes, or endoplasmic reticulum/Golgi²². The APP is continuously metabolized in the central nervous system (CNS), and A β is quickly produced and cleared²³. Several investigations have suggested that the increased production and reduced clearance of A β over a long period might be responsible for the deposition of A β plaques in the brain of AD patients²⁴.

1.2.2 Tau pathology

Tau is the primary microtubule-associated protein (MAP) in a mature neuron in its normal state²⁵. Tau protein is one of the most abundant proteins in neurons, with a role in stabilizing the structure of microtubules²⁶. In AD, tau becomes abnormally hyperphosphorylated and dissociates from microtubules, which causes tau to self-assemble in the form of aggregates into neurofibrillary tangles (NFTs)²⁷ (Figure 1.1). The NFTs can be detected in the CSF at an increased level relative to healthy controls²⁸. Current studies have reported the hyperphosphorylation of thr181 and ser199, suggesting that these specific epitopes have higher diagnostic accuracy in differentiating AD from healthy controls²⁹. In LOAD, the thr231 tau epitope was detected before forming paired helical filaments²⁹, whereas thr181 and ser199 were considered late events in AD³⁰. The hyperphosphorylation of tau further disrupts its interaction with microtubules, kinesin, dynein motor protein function, and axonal transport, which is incompatible with neuronal function and ultimately results in neuronal death.

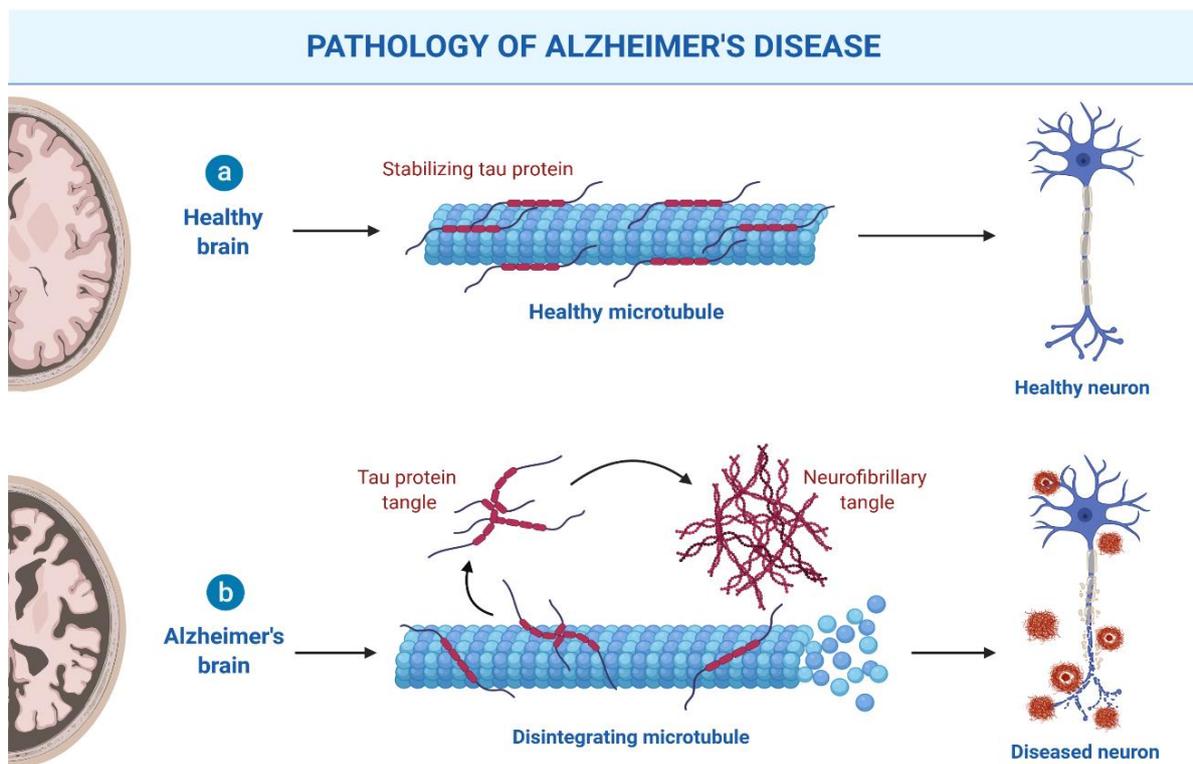


Figure 1.1i: Pathophysiology of Alzheimer's disease. (a) The healthy brain has intact neurons that transmit nerve impulses and interact with adjacent neurons, thereby forming a neuronal network. The axonal microtubules are stabilized by tau, maintaining axonal shape and stability. (b) In AD, tau protein is subjected to various post-translational modifications that reduce the affinity of tau for microtubules and instead, they assemble to form aggregates. Microtubule depolymerization causes loss of axonal integrity, leading to degeneration and eventually cell death, which is anatomically visualized as a shrunken brain in AD cases. This figure was prepared using molecular subcomponents provided in Biorender software (Toronto, Canada).

1.2.3 Treatment strategies

Several failed clinical trials highlight the urgent need to understand the disease before the onset of clinical symptoms, as most treatments have maximum impact and efficiency if implemented at earlier stages of the disease. To date, there is no reliable treatment to cure or halt the progression of the disease. The only approved treatments available are NMDA agonists like memantine and cholinesterase inhibitors (donepezil, galantamine, and rivastigmine), which may give symptomatic relief but do not alter the course of the disease³⁰. Patients with mild to moderate AD are often prescribed cholinesterase inhibitors (inhibitors of enzymes that catalyse acetylcholine breakdown) to increase the half-life and level of acetylcholine in the nervous system. By contrast, memantine is generally prescribed for moderate to severe AD, the mechanism of action involving an alteration of neuronal excitotoxicity by blocking the prolonged calcium ion influx into the postsynaptic terminal.

Current research on AD specific drugs is focused on inhibiting the activity of beta-secretase 1 (BACE1) from reducing the A β burden, increasing clearance of A β (by active or passive immunotherapy), inhibiting GSK3 to reduce the abnormal hyperphosphorylation of tau, or its fibrillation/deposition into NFTs (by active or passive immunotherapy)³¹. Since 2002, when memantine was approved for moderate to severe AD, no new drug has been approved. The only exception is the 2021 accelerated approval of a monoclonal antibody against amyloid β (A β) – aducanumab – by the US Federal Drug Administration (FDA) on the basis that it removes amyloid from the brains of AD patients and potentially leads to cognitive benefit. This approval has been controversial for several reasons, the main reason is that the cognitive benefits remain conclusively established, and the belief that removing amyloid will lead to clinical benefit is not well-supported by evidence³².

Apart from the accumulation of well-defined A β plaques and hyperphosphorylated tau proteins in AD, various biochemical pathways are disrupted in AD, and is, therefore, features a complex web of pathological processes. Other pathological processes involved include neuroinflammation, oxidative stress and synaptic loss, which are potential therapeutic targets.

Recent studies have shown that the accumulation of proteins and other pathological changes in the brain might begin 15-20 years before the onset of clinical symptoms. Therefore, the possibility of disease identification in its early (pre-clinical) stages is also of considerable interest³³. The long pre-clinical trajectory of AD means that neurodegenerative damage can

progress to an irredeemable stage when the patient (or carrier/s) notices symptoms, seeks medical attention, and confidently identifies clinical symptoms diagnosis provided. Furthermore, it is essential to keep in mind the complexity of AD pathology. Drugs targeting a specific molecular pathway have failed several times. Consequently, in this multifactorial disease, multi-target therapies may represent a more effective approach to manage the different pathogenic aspects of AD³⁴.

1.3 Biomarker and Biofluids

A biomarker is an objective indicator either of pathological processes or biological and/or physiological changes in response to an intervention or drug exposure, which can be measured accurately and reproducibly. Biomarker discovery aims to differentiate the disease condition from the corresponding normal control individuals. A secondary goal of biomarker discovery may include a better understanding of pathological pathways, but understanding disease mechanisms is not necessarily a pre-requisite for a disease biomarker.

Various body fluids like CSF, synovial fluid, urine and blood are clinically useful sample types often used for disease diagnosis and biomarker research. In principle, the closer body fluid contact with the diseased organ, the more direct the measure of organ-specific changes. For this reason, blood is often used in pathology testing since many of the organs (e.g., heart, liver, spleen, pancreas, kidneys, and others) of the body come in direct contact with blood and blood vasculature. Blood is also a relatively easy fluid to collect, and venepuncture is a routine and commonly performed procedure, both for clinical purposes and research. By contrast, the CNS is shielded by the blood-brain barrier (BBB), so in this case, CSF comes in closest contact with the brain and spinal cord. The CSF shares some of the protein content from plasma since about 500 mL of CSF is exchanged to blood daily in humans. The CSF is occasionally clinically sampled to assist the diagnosis of some conditions (e.g., meningitis). However, the collection of CSF is invasive and painful, requires expertise to overcome the risk involved, and is generally reserved for specific clinical requirements, and is rarely allowed for purely research purposes.

Body fluids such as urine and saliva are emerging as useful research samples because they are non-invasive, cheap and easy to collect. However, they are generally further removed from the target organ, and the sensitivity and specificity of biomarker detection may be compromised. In saliva, protein composition varies with different physiological conditions, diet, age, gender, and circadian rhythm and sample variations also occur with the procedure of

sample collection and time of collection. Urine is mainly an aqueous solution (greater than 95%) with its content to a significant extent comprised of waste metabolites and electrolytes (like chloride, sodium, potassium), inorganic and organic compounds, and some proteins at low concentrations (<150mg/day), primarily albumin (<30mg/g) in the normal individual. Urine is a valuable sample for urinary system diseases like renal disease, but it is uncertain whether it can be a good sample for other vital organs such as the brain due to low protein concentration. Furthermore, saliva and urine may have microbial contamination as samples are collected externally and maintaining sample integrity for clinical analysis may be more challenging.

By contrast, human blood is ordinarily sterile, relatively easy to collect with minimal training required (venepuncture) and contains analytes from all organs and tissues, including proteins (as leakage markers) such as numerous immunoglobulins circulating in the blood, lipids and metabolites³⁵. The brain, however, has a privileged position because of the BBB, which controls the influx and efflux of biological substances between the brain and the blood, and helps control the microenvironment of the brain. The BBB comprises the endothelial cells of the capillaries, pericytes and the foot processes of astrocytes³⁶. Transport across the BBB occurs by passive and active diffusion, carrier-mediated transport, receptor-mediated transport, adsorptive mediated transport and cells diapedesis. The BBB is relevant to the study of blood biomarkers of AD in several ways. First, A β is reversely transported from the brain to the blood by binding to LRP1, resulting in its clearance from the brain. There is evidence that disruption of the BBB occurs early during AD. This is due to immune cell activation and migration and the release of cytokines that affect BBB integrity. Several factors, including vascular factors and *APOE ϵ 4*, has been shown to cause BBB disruption. The presence of blood biomarkers of AD is thereby facilitated by the early changes in BBB integrity related to the disease³⁷.

In terms of pathways association, there are a number of significant genetic variables linked with AD, however even these cannot completely account for the entirety of AD cases. Rather than focusing on specific genes, a more appropriate approach would be to investigate AD as a result of disruptions to whole biological networks. The wider pathway that plays a significant role in AD pathogenesis is shown in figure 1.2. Numerous mechanisms have been proposed to contribute to AD, including neuroinflammation, oxidative stress, defects in mitochondrial dynamics and function, cholesterol and fatty acid metabolism, as well as impairments in glucose energetic pathways in the brain, autophagy failure, and other less studied mechanisms. I have included a short overview of these processes in this paragraph. Section 1.9 contains

additional detailed information. While the different processes are explained individually, they are intricately linked and often act synergistically in the CNS.

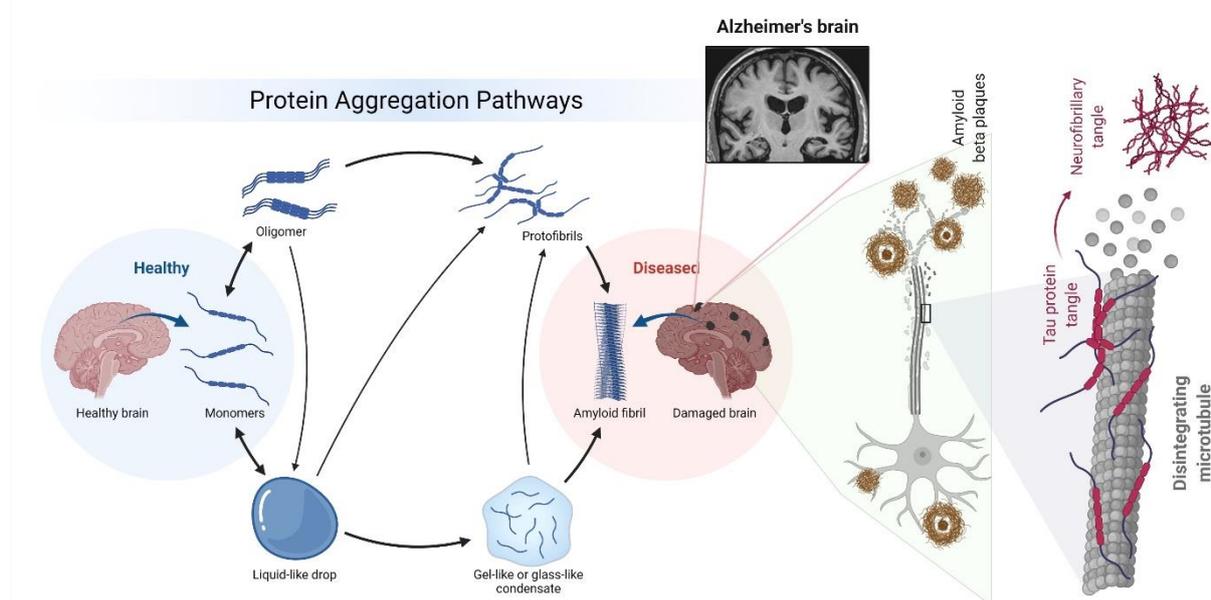


Figure 1.1ii Protein aggregation mechanisms in AD are shown schematically. Two portions are shown in the illustration. The left half outlines the progression from a healthy to a diseased state, while the right section contains material specific to AD pathology.

In regards to biomarkers, blood, plasma, and serum are exceptional sample types to study disease biomarkers, and in particular, the protein component of blood is one of the most complexes of all tissues. However, the high dynamic range and complexity of plasma protein abundance also present a challenge for analysis, particularly when using non-targeted discovery approaches, such as mass spectrometry-based proteomics. To address this challenge, plasma samples need to be extensively processed to analyse a significant part of the plasma proteome. Low abundance proteins are also frequently those which originate from organs and are therefore most relevant as potential disease biomarkers or biomarkers of physiological change. Various methods are available to enrich low abundance proteins at protein or peptide levels, including immunoaffinity-based depletion of high abundance proteins, bead-based enrichment, and offline chromatography-based separation approaches³⁸.

1.4 Methods for identification of plasma proteins:

1.4.1 ELISA (enzyme-linked immunosorbent assay)

ELISA is one of the most widely used methods for quantification and characterization of specific analytes using antigen-antibody interactions on the surface of a microtiter plate³⁹. Advantages of the ELISA assay include: (1) it is a safe, easy and cost-effective procedure, reagents are relatively low cost, and the procedure is simple to perform; (2) it has high efficiency, as samples do not require complicated pre-treatments before the analysis; and (3) it has high sensitivity and specificity because of specific antigen-antibody interactions. Conversely, disadvantages of ELISA assay include: (1) the synthesis of antibodies is expensive as they require high-quality cell culture media to obtain a specific antibody; (2) antibodies are unstable and require optimum refrigeration for storage and transportation; (3) insufficient blocking of immobilized antigen surface of a microtiter plate can increase the possibility of false-positive results; (4) antibodies are not readily available for several analytes⁴⁰, and (5) lack of sufficient antibody specificity can result in an increased risk of false-positive results.

1.4.2 Western blot (WB)

Western blotting is one of the most common procedures in biochemical labs. It is also an antibody-based approach and therefore has similar advantages and limitations as described above for ELISA assays. For Western blotting, proteins are first separated by size (1D or 2D SDS/PAGE), and then transferred to a membrane such as nitrocellulose or PVDF by contact transfer or electroblotting. The membrane is exposed to the specific primary antibodies which bind to the specific protein/epitope, if it is present. Following the primary antibody incubation step, the membrane is exposed to a secondary antibody, which binds to any primary antibody which may be present, and provides the detection signal (such as optical or chemiluminescent development)⁴¹. This approach is beneficial not only in research but also in medical or diagnostic labs; tests for both HIV and Lyme disease, for example, involve an enzyme-linked immunosorbent assay (ELISA) test, followed by a Western blot if the ELISA tests are positive. Western blots are often used to determine relative protein levels between samples. They also establish the molecular weight of the target, which may provide insight into its post-translational processing. Furthermore, they can provide additional evidence of primary antibody specificity for the target protein of interest, especially if coupled with use of positive and negative control samples. Proteins from tissue/cell lysates are separated by gel electrophoresis according to their molecular weight.

Advantages of Western blotting are: (1) it is a safe and eco-friendly procedure; (2) gel electrophoresis sorts proteins of different size, charge and conformation, which gives an idea of the size of protein or polypeptide of interest; and (3) specificity of antigen-antibody interaction can identify the protein of interest in a mixture of thousands of

proteins. Disadvantages of Western blotting include: (1) it is non-quantitative as Western blot can only determine the presence/absence of a specific protein but cannot quantify the exact amount of the protein present in the sample, since standards are generally not used; (2) as with ELISA assay, Western blotting requires antibodies which are expensive, unavailable for several proteins, and often cross-react with multiple similar epitopes, and may therefore not have sufficient specificity for the target protein; (3) the protein sequence epitope against which an antibody is raised may be masked if the target protein is post-translationally modified (e.g., by glycosylation, which can be reasonably extensive); (4) insufficient antibody specificity is a common problem, since the epitopes of commercial antibodies are often not fully characterised and consequently false positive or negative results may occur⁴².

1.4.3 Immunohistochemistry (IHC)

Immunohistochemistry (IHC) is an important application of monoclonal and polyclonal antibodies to determine the tissue distribution of an antigen of interest in health and disease. Immunohistochemistry (IHC) combines anatomical, immunological and biochemical techniques to image discrete components in tissues using appropriately labelled antibodies to bind specifically to their target antigens *in situ*⁴³. IHC makes it possible to visualize and document the high-resolution distribution and localization of specific cellular components within cells and within their proper histological context. While there are multiple approaches and permutations in IHC methodology, all the steps involved are separated into two groups: sample preparation and sample staining. As outlined in the ELISA and Western blotting sections above, the approach has similar advantages and limitations to other antibody-based techniques.

1.4.4 Evolution of mass spectrometry:

In 1911, Joseph John Thomson developed the first mass spectrometer, a parabola spectrograph used to analyse the isotopes of neon after the pioneering work on electrons in 1899. He was later awarded a Nobel Prize in Physics "in recognition of the great merits of his theoretical and experimental investigations on the conduction of electricity by gases"⁴⁴. Thomson's student Francis W. Aston further developed the instrument and played a significant role in improving the resolving power to study other isotopes of non-radioactive elements, and was awarded the Nobel Prize in Chemistry in 1922. Promising developments in areas such as increased sensitivity, resolving power and ionization sources in the instrumentation led to the use of mass spectrometry in a wide range of fields, including chemistry, physics and more recently, biology and medicine, for characterization and quantification of biomolecules from a variety of

complex samples, demonstrating the versatility of mass spectrometry for interdisciplinary research ⁴⁵.

1.4.4.1 What is mass spectrometry?

Mass spectrometry is a technique used to determine the mass-to-charge ratio (m/z) of gas-phase ions. Mass spectrometry is frequently coupled with a chromatographic technique such as gas or liquid chromatography to separate complex sample mixtures and provide multidimensional information, such as chromatographic retention time, parent ion mass, and tandem instruments fragment masses to provide structural information. Some excellent books and review articles have been written outlining mass spectrometry principles and applications, particularly regarding biological, biomedical and proteomics applications⁴⁶⁻⁴⁸.

1.4.4.2 Mass spectrometry-based proteomics:

The term proteomics was first introduced into the scientific lexicon by Wasinger et al. 1994 at Macquarie University⁴⁹, Australia, though the approach was used at least a decade before the term was coined. The approach depends on three main components: (1) mass spectrometry for high throughput processing of proteins or protein mixtures; (2) software/algorithms which can convert mass spectrometric data of peptides into amino acid sequences (e.g., Mascot, Sequest, Andromeda, and others); and (3) availability of comprehensive protein databases (derived from genome sequencing, now available for many organisms). Proteomics is a powerful tool for identifying and quantifying all the proteins in an organism expressed by a genome under a defined environmental condition such as pathophysiological processes, response to environmental stimuli, treatment responses in patients, and others. The study of the constantly evolving proteome is essential for understanding complex biological machinery, its normal function and how it changes in response to specific environmental variables, or as a function of disease processes. For example, one human gene can result in more than ten protein variants, and changes in the level and pattern of post-translational modifications can change the differential expression and functioning.

Several other challenges in identifying the changing proteome include the complexity of human fluids, tissues, and cells. Multidimensional Protein Identification Technology (MudPIT) approaches are commonly used as proteomics tools. However, sample complexity can cause signal suppression, and therefore limit the full potential range of the identified proteome. To improve proteome coverage, sample fractionation can help, and several separation technologies can be used, including two-dimensional polyacrylamide gel electrophoresis (2D-PAGE)⁵⁰ and nano liquid chromatography (nanoLC) methods, in addition to improvement of ionization

methods or scan rates in mass spectrometry such as matrix-assisted laser desorption/ionization (MALDI)⁵¹ and electrospray ionization (ESI)⁵².

Furthermore, advances in instrument design have improved the sensitivity and mass resolving power of mass spectrometry instruments, including Fourier transform ion cyclotron resonance and orbitrap technology. This has vastly improved the identification and quantification of hundreds to thousands of proteins in a short period.

1.4.4.3 Proteomic workflow (MudPIT approach):

In discovery and targeted proteomics experiments, proteins are extracted from the sample and cleaved into defined peptides using a protease, typically trypsin. The digested peptides are fractionated using liquid chromatography and analysed by electrospray ionization coupled to mass spectrometry⁵³. The mass spectrometer separates charged ions in its analyser, and the ions are then detected and reported as the mass-to-charge (m/z) ratio. With a discovery-based strategy, the aim is to identify protein expression changes, in addition to providing protein specificity which results from peptide sequencing by MSMS. Such an approach is also called data-driven proteomics and is an observational approach that often precedes and underpins subsequent hypotheses. In this approach, all peptide ions which fall above a specified signal intensity threshold are transmitted to the mass spectrometer first-stage analyser (parent ions), followed by a second stage in which the parent ions are fragmented, producing daughter ions (also called MSMS fragment ions). Bioinformatics software uses the parent and related daughter ions to interrogate protein sequence databases (e.g., SwissProt, NCBI, and others) in a pattern recognition approach to identify proteins that best match the available protein sequence data. Discovery-based proteomics experiments often report the identification of 100s – 1000s of proteins, and they also provide some rudimentary quantitative information, such as fold change or relative quantification, based on techniques such as spectral counting or peak area integration. However, they do not provide absolute quantification of concentration. By contrast, targeted proteomics experiments are designed to detect specific peptide ions resulting from proteins of interest. Their selection generally requires some prior knowledge, this approach is frequently used in hypothesis-driven work, and involves either single reaction monitoring (SRM) or multiple reaction monitoring (MRM), which have high sensitivity, reproducibility, and quantitative accuracy⁵⁴. Figure 1.2 shows typical workflows for label-free discovery-based mass spectrometry and targeted mass spectrometry.

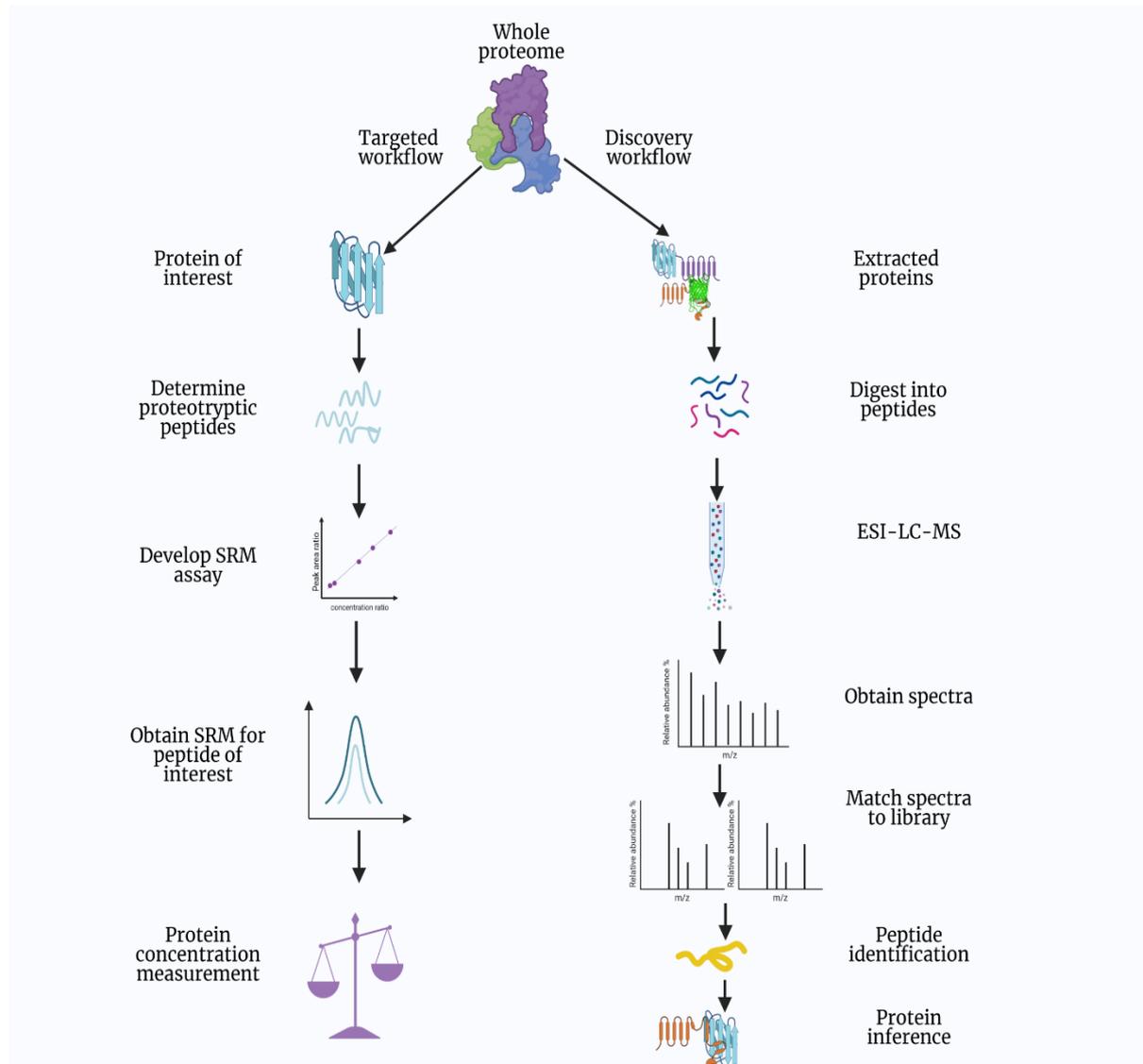


Figure 1.2: Workflows of discovery-based and targeted proteomics using mass spectrometry. This figure was prepared by Ms Kaur, using molecular subcomponents provided in Biorender software (Toronto, Canada).

1.4.4.4 Quantitation methods of MS

After initial protein identification, quantification of the identified proteins is a more challenging task in MS-based proteomics. However, several MS-based quantification methods have been developed for both relative and absolute measurements (Figure 1.3), and proteomics technologies continue to develop better and more efficient quantification approaches.

The digested peptides derived from two or more samples to be compared are separated by LC-MS and then analysed by MS and MS/MS (tandem mass spectrometry) followed by database searching of the raw data, to identify the protein sequences present. One commonly used label-

free quantification approach, is that identical proteins derived from two different biological samples or states are quantified based on either the relative intensity of the precursor mass or else a fragment ion mass from the MS/MS spectrum⁵⁵. Label-free quantification is a relatively simple and cost-effective approach, with minimal sample processing, since it does not involve sample labelling with mass tags or stable isotopes (such as iTRAQ, SILAC, TMT, etc). Using a label-free MudPit MS-based approach, relative quantification of proteins in two or more biological samples is typically achieved using either of two strategies: (1) spectral counting, and/or (2) extracted ion chromatograms (XICs)/spectral signal intensity/peak intensity⁵⁶. Spectral counting is a relative protein quantification approach which “counts” the number of times all peptides corresponding to a particular protein were sequenced. The more abundant protein will result in a greater opportunity for ionisation of the peptides which comprise its sequence, detection above the limit-of-detection threshold, and therefore a greater number of LC MSMS spectral counts. Normalization of spectral counts is required by correcting for the protein length (exponentially modified protein abundance index, emPAI) or the spectral counts for each protein by the mass or protein length (normalized spectral abundance factor, or NSAF)⁵⁷. Another common approach for label-free quantification is by use of peak intensity (either peak height or peak area). Proteomic quantification of this kind is analytically demanding, and is typically facilitated with use of software, such as MaxQuant, ProteomeDiscoverer, Progenesis, etc. Such software typically utilises a number of variables available in the LCMSMS data, such as quantification of peptide ion intensity, m/z , and chromatographic elution time to ensure high analytes specificity. Such software typically also performs some basic statistical analyses for across sample comparisons. In this method, the intensity and number of selected precursor ions at a particular m/z are summed, and the peak areas are used to measure relative abundance across samples. Some software has additional refinements, with the aim of improving quantitative accuracy. For example, in the intensity-based absolute quantification (iBAQ)⁵⁸ method, the sum of XIC peptide intensities is divided by the number of theoretically observable peptides.

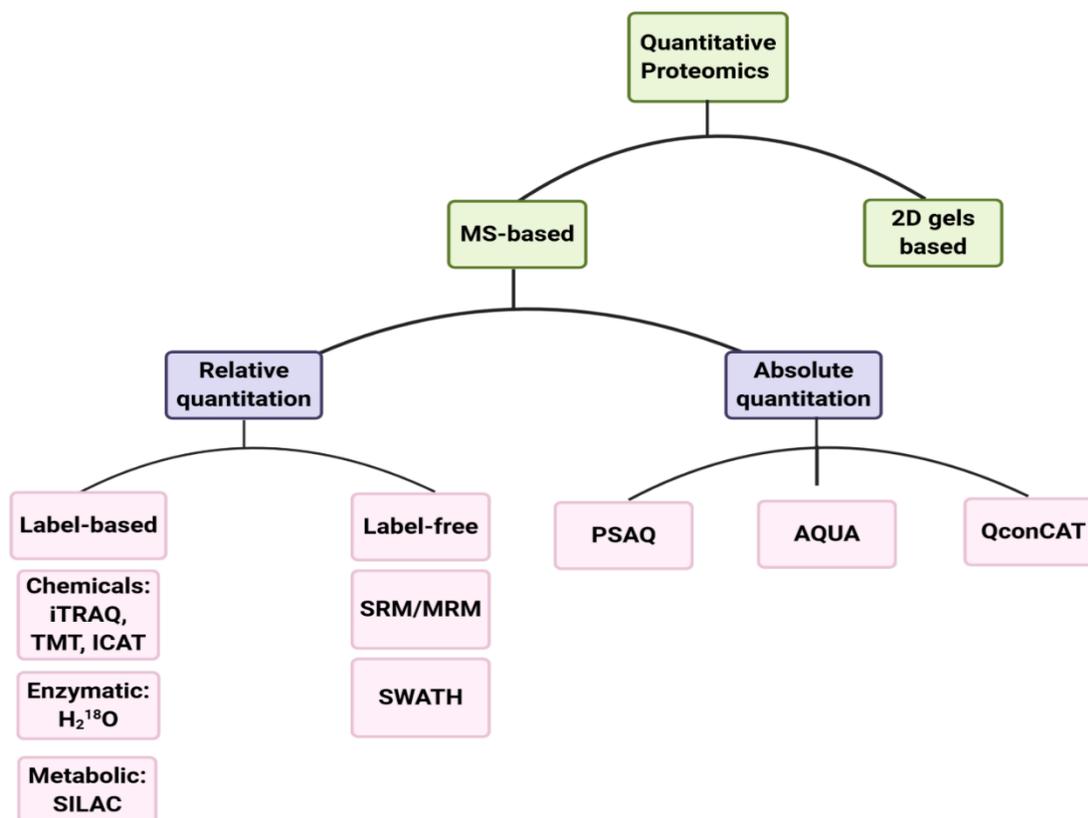


Figure 1.3: An overview of MS-based quantitative proteomics: MS-based quantification can be achieved either by relative or absolute quantification. Both relative and absolute quantification can be classified into labelled and label-free quantification, each performed by several approaches. Relative quantification techniques: Isobaric tags for relative and absolute quantitation (iTRAQ), tandem mass tag (TMT), Isotope-coded affinity tag (ICAT), Selective/Multiple Reaction Monitoring (SRM/MRM), Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH). Absolute quantification techniques: Protein Standard Absolute Quantification (PSAQ), Absolute Quantification (AQUA), Quantification conCATemer (QconCAT)

The ability to differentiate stable isotopes based on small differences in mass is a capability unique to mass spectrometry, and enables a powerful quantitative tool, which is the use of structurally identical internal standards. There are several ways in which stable isotopes can be utilised in mass spectrometry. One of the most common approaches is stable isotope dilution mass spectrometry, in which a labelled analyte (e.g., peptide) is used at a constant amount across all samples (including calibration standards and controls). It is therefore one of the most robust sample normalisation techniques, since both analytes and its internal standard are not only structurally identical, but they are analysed within the same LCMSMS run, thereby eliminating run to run variation. If used at the earliest feasible step at the beginning of sample processing, a stable isotope labelled internal standard can correct for variations in sample extraction efficiency, sample handling losses during processing, analyte instability during processing, variation in sample delivery into the mass spectrometer, and any other variables affecting sample preparation. As with label-free proteomics methods, use of stable-isotope

labelling, coupled with LC-MS/MS, provides a fast method to simultaneously provide quantitative and qualitative information (protein identification) in a single run, with the added advantage of a powerful normalisation approach enabled by the use of the stable-isotope labelled peptides. Use of specific isotope labelled analytes/peptides, is possible in targeted approaches, but stable isotopes can also facilitate discovery-based proteomics. A variety of methods are available for labelling of complex sample mixtures, enabling analysis of multiple samples within a single run. The labelling methods are usually categorized into three types: chemical labelling, enzymatic labelling, and metabolic labelling. Isotopic elements can be incorporated into the sample at the peptide level (enzymatic labelling), protein level (chemical and metabolic labelling), or after enzymatic digestion (chemical labelling). Isobaric tags for relative and absolute quantitation such as Isotope-Coded Affinity Tags (ICAT)⁵⁹, isobaric Tag for Relative and Absolute Quantitation (iTRAQ), and Tandem Mass Tags (TMT) can compare multiple samples (multiplex) combined in one LC-MS run⁶⁰ (Figure 1.5). In metabolic labelling, stable isotope labelling by amino acids in cell culture (SILAC) and enzymatic labelling such as H₂O/H₂¹⁸O labelling⁶¹ (Figure 1.4) are benchmark labelling methods for quantitative proteomics.

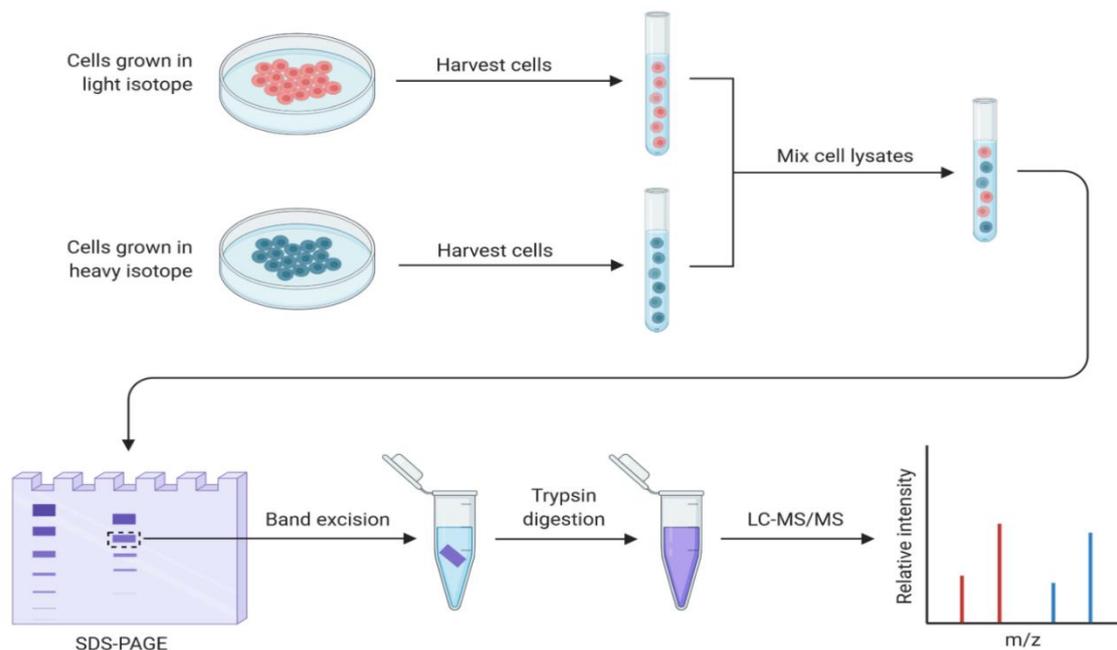


Figure 1.4: An overview of the SILAC labelling protocol. In a conventional SILAC experiment, cells are grown in light (red) and heavy (green) SILAC media until the cell grown in heavy media have fully incorporated the heavy amino acids. Labelled cultures are mixed upon the treatment, and the relative protein changes are determined from the ratio of heavy and light peptide signals in MS (This figure was generated using Biorender software (Toronto, Canada).

Table 1.2: This table, adopted from Ludwig et al., 2018⁶² shows the advantages and limitations of SWATH-MS compared to data-dependent (DDA) and targeted (SRM, PRM) proteomics.

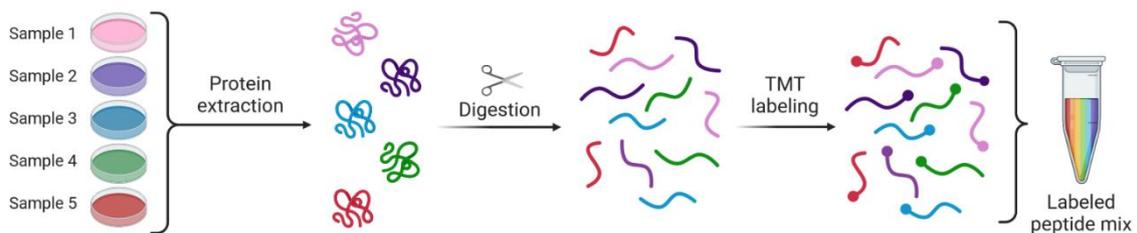
	Data independent acquisition-based SWATH-MS	Data-dependent acquisition (DDA)	Targeted acquisition (SRM and PRM)
Ease of data acquisition	** Easy, requires definition of mass range to cover, precursor isolation window width and number of MS2 scans per cycle	*** Easiest, default setup on most mass spectrometers, requires definition of TopN method, MS2 trigger threshold and dynamic exclusion time	* Hardest, requires generation and optimization of targeted peptide assays for data acquisition
Ease of data analysis	* Currently hardest, requires peptide query parameters, sophisticated software tools, large informatics resources	*** Currently easiest, multitude of pipelines available	** Currently easy, several software tools developed for manual and automated analysis
Breadth of protein and peptide detection/multiplexing	*** 10,000s of peptides per MS injection quantifiable	*** 10,000s of peptides per MS injection quantifiable	* 10s - 100s peptides per MS injection quantifiable
Selectivity/sensitivity/dynamic quantification range	** 4 orders of magnitude per MS injection	** 4 orders of magnitude per MS injection	*** 4-5 orders of magnitude per MS injection
Reproducibility/data consistency	*** High, due to peptide-centric scoring analysis	* Low, due to stochastic sampling in DDA	*** High, due to targeted data acquisition
Retrospective targeting (using chromatogram extraction)	*** Possible on MS1 and MS2 level	** Possible on MS1 level only	* Not possible due to targeted data acquisition

*Least optimal performance.

**Medium performance.

***Best performance.

1) TMT labeling protocol



2) Data collection and analysis

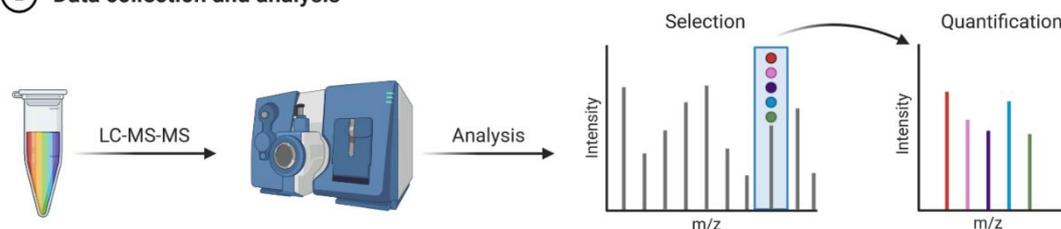


Figure 1.5: An overview of TMT labelling: 1) After lysis of cultured cells, cysteine bonds are reduced and alkylated. Next, extracted proteins were digested using trypsin. The resulting peptides were labelled with TMT, and all samples were mixed to perform a single run. 2) The TMT labelled peptide mix is run on LC-MSMS, and further data are collected and analysed using bioinformatics tools to mine meaningful biological information from the protein data.

Advantages and disadvantages:

Protein quantification using labelling strategies often requires pairwise comparisons, which becomes a limitation when quantifying proteins that only exist in a given physiological state. Lack of retrospective quantification and high costs are major disadvantages of labelled approaches. On the other hand, one of the main advantages of label-free protein quantification is that the number of samples per experiment are not constrained by numbers of labels available e.g., iTRAQ is limited to 8-plex format, TMT is limited to 10-plex format, SILAC similarly restricts experimental design to numbers of available labelled amino acids, while the numbers of samples per experiment are unlimited using the label-free approach. In addition, the analysis can be conducted retrospectively, allowing for protein quantification/abundance across a longitudinal set of studies. Furthermore, labels generally represent a substantial experimental cost in terms of both reagents and time involved in sample preparation. The analysis can be conducted retrospectively, and with greater flexibility, since data is collected on samples individually, rather than pooled samples, allowing for protein quantification across a longitudinal set of studies, or reanalysis of data using a different comparison approach.

1.5 Biomarkers of AD pathology:

In 2018, the National Institute on Aging and Alzheimer's Association (NIA-AA) research framework presented the possibilities of the addition of more biomarkers to the ATN framework for AD [presence of β -amyloid (CSF A β or amyloid PET: "A"), hyperphosphorylated tau (CSF p-tau or tau PET: "T"), and neurodegeneration (atrophy on structural MRI, FDG PET, or CSF total tau: "N")]⁶³. They introduced the ATNZ classification, where Z could characterise the underlying pathophysiological changes. However, except for A β and tau biomarkers, other AD markers such as inflammation often lack specificity, and new ones are being developed at a relatively slow rate. Keeping in mind the urgency of developing strategies to combat the rising number of AD cases, it would be more effective to track the disease in its earlier stages, with a focus on stopping or slowing disease progression rather than focusing on stages post-clinical diagnosis of AD, at which point disease progress has likely caused life-style altering damage to the brain.

1.5.1 CSF biomarkers of AD:

Numerous research has evaluated three main CSF biomarkers for Alzheimer's disease ($A\beta_{42}$, T-tau, and P-tau), and several more Alzheimer's disease markers are emerging in the literature²¹. $A\beta$ peptide is a hallmark of AD, whose complex lifecycle not only produces a plethora of chain length variants such as $A\beta_{38}$, $A\beta_{40}$, $A\beta_{42}$, $A\beta_{43}$ ¹⁶ but also multimeric aggregate structures such as oligomers, fibrils and plaque¹⁷. Of these, the 42-amino-acid peptide ($A\beta_{42}$) is the most studied variant due to its enhanced tendency to aggregate⁶⁴. Tau is one of the most abundant proteins in neurons, with a role in stabilizing the structure of microtubules²⁶. In AD, tau is hyperphosphorylated, dissociated from microtubules, and detected in CSF at an increased level relative to healthy controls²⁸. Apart from the well-studied CSF biomarkers, loss of integrity of white matter pathways that consist of myelinated axons rich in neurofilaments (light, medium and heavy) in AD patients has led to biomarker discovery. Several studies have demonstrated a significant increase in the concentration of neurofilament light (NfL) in AD patients' CSF compared to healthy controls⁶⁵. An increase in NfL concentration has also been reported in other neurodegenerative diseases⁶⁵, thereby limiting the diagnostic specificity of CSF NfL for the differentiation of AD from other neurodegenerative diseases. However, it may serve well as a marker of neurodegeneration. Another emerging study showed the disruption and increased permeability of the blood-brain barrier (BBB) correlated with an increased level of platelet-derived growth factor receptor- β (sPDGFR β) in the hippocampus⁶⁶. The sPDGFR β is highly expressed in brain capillary pericytes, vascular mural cells, and arterial vascular smooth muscle cells (VSMCs). Neurogranin (NRGN) is a calmodulin-binding protein secreted by neuronal cells that are highly expressed in the cortex, hippocampus, and amygdala, and it is mainly concentrated at the dendritic spines⁶⁷. The level of NRGN is elevated in both mild cognitive impairment (MCI) and dementia due to AD compared to cognitively normal elderly subjects⁶⁸. Synaptic degeneration is an early event in AD, making synaptic biomarkers relevant for early diagnosis. A recent meta-analysis, which included 16 independent studies, confirmed CSF NRGN as a potential marker for AD. The authors recommended that NRGN be added to the panel of existing biomarkers to improve diagnostic accuracy⁶⁹.

Neuroinflammation is a well-known dysregulated mechanism, which may cause pathophysiological changes in neurodegenerative diseases, including AD. Microglial activation in the brain is the primary innate immune response to pathogens and amyloid plaques in AD. Microglia are involved in the reduction in $A\beta_{42}$ neurotoxicity through phagocytosis of

amyloid plaques to maintain brain homeostasis⁷⁰. Activation of various immune system proteins includes increased Triggering Receptor Expressed on Myeloid cells 2 (TREM2) in CSF. The TREM2 protein is a receptor expressed on the surface of microglial cells⁷¹. As a multi-functional protein, TREM2 regulates phagocytosis, inhibiting the proinflammatory response and removing apoptotic neurons^{72,73}. The soluble form of TREM2 (sTREM2) has been detected in CSF and may result from proteolytic cleavage and release into the interstitial fluid of membrane-bound TREM2⁷⁴⁻⁷⁶. An increase of CSF sTREM2 has been reported as an early biomarker in AD patients^{11,77-80}. A higher CSF sTREM2 at a given biomarker level of A β and pathologic tau were associated with larger grey matter volume⁸¹ and slower subsequent cognitive and clinical decline in symptomatic elderly participants⁸². In the symptomatic phase of AD, these results support a protective role of elevated sTREM2 in AD. However, it is unclear whether increased CSF sTREM2 is associated with the reduced longitudinal increase of fibrillar A β ⁸³. A rare homozygous loss-of-function mutation in the human TREM2 gene can cause an acute form of FTD-like dementia linked with cystic bone lesions known as Nasu-Hakola disease^{84,85}. Interestingly, heterozygous missense mutations in the TREM2 gene have been recently described to significantly increase the risk of AD, with an odds ratio similar to that of carrying an *APOE* ϵ 4 allele and other neurodegenerative diseases, including PD, FTD, and ALS^{86,87}. The *TREM2* mutation might be responsible for increased density of neurofibrillary tangles and amyloid plaques, upregulation of proinflammatory cytokine levels, and downregulation of protective markers⁸⁸.

Chitinase-3-like protein 1 (CHI3L1), also known as YKL-40 and HCgp-39, was proposed as a neuroinflammatory biomarker⁸⁹. This protein is generally produced in reactive astrocytes, and reports show an increased concentration of CSF YKL-40 in AD and cognitively normal late-middle-aged individuals⁹⁰⁻⁹³. An increased concentration of YKL-40 was reported in MCI patients carrying an *APOE* ϵ 4 allele compared to noncarrier MCI patients. These studies suggest that YKL-40 may be a possible inflammatory biomarker for identifying individuals converting from MCI to AD⁹⁴. On the contrary, some studies reported no difference in YKL-40 levels between *APOE* ϵ 4 carriers and noncarriers⁹¹. More extensive cohort studies are needed to understand the correlation between CSF YKL-40 levels vs. *APOE* ϵ 4 and progression from prodromal MCI to AD and whether YKL-40 combined with core CSF biomarkers could be used as a risk predictor⁹⁵. Additionally, the CSF levels of pre/postsynaptic proteins GAP43,

SNAP25, and SYT1 were able to distinguish AD, and MCI from other non-AD dementia groups and control groups requires further attention⁹⁶.

1.5.2 Blood biomarkers of AD:

The main blood-based biomarkers, i.e. $A\beta_{42/40}$ ratio, tTau and pTau (181, 217 and 231 isoforms) and NfL, have been widely studied and might approve for clinical use in the near future⁹⁷⁻⁹⁹. Several studies have shown a significantly lower concentration of plasma $A\beta_{42/40}$ found in AD compared to control samples¹⁰⁰. Interestingly, immunoprecipitation and mass spectrometry studies identified the APP 669–711/ $A\beta_{42}$ and $A\beta_{40}/A\beta_{42}$ ratios significantly correlated with levels of $A\beta_{42}$ in the CSF and amyloid PET imaging¹⁰¹. A longitudinal study of the ADNI cohort found plasma NfL level to be increased around 7 years before the onset of clinical symptoms in individuals carrying familial AD mutations, and that NfL levels correlated with change in cognitive scores, imaging measures of hypometabolism and neurodegeneration¹⁰². However, NfL is considered a marker of neurodegeneration and not specific to any particular neurodegenerative disease, showing poor specificity for AD^{101,103}. Additionally, β -secretase 1 (BACE1) is an enzyme responsible for cleavage and generation of $A\beta$ peptides from APP¹⁰⁴ and is categorized as an emerging blood-based biomarker¹⁰⁵. ELISA based assays depicted increased BACE1 activity in MCI or AD patients as compared to healthy controls^{106,107}. A longitudinal study over a 3-year follow-up period reported significantly higher plasma BACE1 activity in MCI, which progressed to AD, than individuals who remained cognitively normal over the 3-year follow-up period. In addition to the CSF mentioned above biomarker, plasma BACE1 activity can be used as a potential biomarker to predict the progression from MCI to AD¹⁰⁶.

1.5.3 Imaging Biomarkers of AD:

Advances in the field of neuroimaging have resulted in several imaging biomarkers for neurodegenerative diseases. Neuroimaging offers flexibility in terms of targeting distinct pathophysiological and age-related mechanisms such as pathological aggregates (amyloid and tau positron emission tomography - PET), functional decline (functional magnetic resonance imaging - fMRI – activity, network correlations), structural decline (MRI based volumetry, cortical thinning), connectivity decline (MRI diffusion anisotropy). One of the major challenges in imaging is that PET tracers require high binding affinity and the ability to cross the blood-brain barrier without being metabolized¹⁰⁸. Amyloid tracers are the most well developed, and the first one to be developed was the C11-PiB (Pittsburgh B compound). There

are now three F18 tracers available— 18F-florbetapir, 18F-florbetaben, and 18F-flutemetamol^{109,110}. A recent study showed that tau ligands to visualize NFTs strongly correlate with the pattern of neurodegeneration and cognitive impairment in AD compared to imaging amyloid deposits¹¹¹.

Furthermore, first-generation PET tracers AV1451 or T-807 were able to reproduce the features of Braak histopathological stages¹¹². These PET tracers were limited by their off-target binding of monoamine oxidase A/B, and increased striatal retention, which was inconsistent with autopsy reports. Most recent tracers have overcome this off-label binding, but more clinical studies are needed¹¹³. Second-generation tau ligands such as [18F] PI-2620 and [18F] MK-6240 were developed to avoid off-target binding and improve diagnostic accuracy¹¹⁴. Besides this, several innovative imaging techniques/methods are under development, such as translocator protein (TSPO)-PET to assess the microglial/neuroinflammation activation¹¹⁵, synaptic density/loss and epigenetic modifications¹¹⁶. Neuroimaging biomarkers have several limitations: (1) they are expensive and therefore unlikely to be useful for population-based screening or longitudinal studies requiring repeat measures; (2) they are invasive, in that tracer dyes need to be infused, and require highly-trained specialists to implement; (3) several studies show that brain plaque levels do not correlate well with disease severity. Other studies show that substantial plaque levels are also present in some individuals who do not meet the criteria for dementia diagnosis (i.e., “normal” individuals). Therefore, the relationship between pathology and disease severity is not always clear. Current thinking suggests that soluble A β oligomers may be the toxic variant, but these do not appear in PET imaging¹¹⁷.

1.5.4 Additional biomarkers of AD:

Clinically relevant biomarkers for Alzheimer's disease continue to be in great demand. However, owing to the limitation of currently available biomarkers (as discussed above), other biomarker approaches are being studied. In particular, a biomarker within a low or non-invasive tissue sample would be highly beneficial.

Some compounds are directly expressed in the salivary glands, while others pass from the blood via active transport and passive diffusion. Several studies show that promising biomarkers are secreted in saliva, including A β ₄₂, p-tau, t-tau, acetylcholinesterase, and lactoferrin¹¹⁸. A two-fold increase in the level of A β ₄₂ was detected in saliva from AD cases compared to controls¹¹⁹⁻¹²¹. By contrast, another study reported that significant findings were confined to mild AD

cases and were not observed in severe AD¹²¹. However, other studies have reported no changes in AD vs control saliva using similar antibody-based methodologies¹²². The discrepancies in data can be due to several reasons. The source of the biomarkers excreted into the saliva remains unknown. To determine the biomarker's concentration reference intervals, more significant sample numbers must be examined. Another factor to examine is the biomarker's diurnal variation depending on when it was collected (close to a meal/drink or not), time of day (i.e., circadian variability), manner of collection, and others. The diminished capacity of AD patients to self-care raises the question of whether poor dental health or cleanliness could impair biomarker detection. It is therefore critical to standardise sampling, processing, and analytical processes in order to assess the reproducibility of the results^{118,123}. In the case of tau protein, similar issues were reported with identification of an increase in T-tau/P-tau181 in some studies while others demonstrated no change of salivary T-tau using a Luminex assay¹²⁴. Lau et al., 2015 demonstrated no change in salivary T-tau but increased P-tau¹²⁵. Published data on salivary tau is highly inconsistent, making it an unreliable marker. Microbial peptides such as lactoferrin, secreted in saliva, are positively correlated with CSF A β ₄₂ and Mini-Mental State Examination (MMSE) scores in AD cases and negatively correlated with T-tau protein¹²³. In a longitudinal study, a group of control participants presented with low concentrations of lactoferrin (< 7.43 μ g/mL), and 78% of them converted to AD or MCI within 5 years. Interestingly, control individuals with an average concentration of lactoferrin (> 7.43 μ g/mL) did not convert within the same time frame. Another potential salivary marker is acetylcholine (AChE) which decreased with age and was downregulated in AD patients compared to age-matched controls^{123,125}.

1.6 Proteomics-based emerging biomarkers:

In the case of neurodegenerative diseases, replacing established and specific markers of PET imaging and/or CSF biomarkers with blood-based biomarkers is questioned¹²⁶. The main advantages of blood-based biomarkers include; (1) that they can be more frequently and easily obtained in the context of community-based studies, large cohorts, or tracking changes in response to drugs or lifestyle changes. (2) blood is one of the most common sample types and is widely used in pathology testing. It involves a relatively low-risk type of sampling, even in the elderly, and requires just a moderate level of training, specifically venepuncture.

Proteomics technologies allow unbiased hypothesis-free biomarker discovery and have the potential to identify several candidate markers relatively quickly. In a recent systematic review and meta-analysis, Rehiman et al., 2020 identified 48 candidate proteomics biomarkers of MCI and AD (out of a total of 207 proteins identified), reported in > 2 studies. This review highlighted six potential blood-based markers; apolipoprotein A-1 (ApoA-1), insulin growth factor binding protein-2 (IGFBP-2), alpha-2-macroglobulin (α 2M), fibrinogen- γ -chain, afamin, and pancreatic polypeptide (PP), which had similar trends in >3 independent cohorts¹²⁷. Several studies suggest that complement activation in the central nervous system (CNS) might contribute to A β aggregation and microglial activation, as A β plaques were detected in the presence of complement proteins such as C1q, C3 and C4¹²⁸. Other than complement components, α 2M has been reported to bind to A β with high affinity assisting its degradation and clearance. In 2019, Hall and colleagues reported a significant correlation between the upregulation of α 2M with cognitive decline¹²⁹ and α 2M has previously been reported as an AD risk factor^{130,131}. Three independent cohorts found another interesting marker named afamin. Afamin, mainly secreted by the liver to transport vitamin E, showed a consistent downregulatory trend in plasma samples of AD vs healthy controls. *In vitro* studies have shown that afamin facilitates transport of α -tocopherol across the blood-brain barrier¹³². In the brain, afamin may function as an antioxidant, and studies suggest that downregulation of afamin might lead to neurodegeneration in AD brain¹³³. Another potential biomarker reported in the Rehiman et al. 2020 meta-analysis review was ApoA-1, mainly localises to high-density lipoproteins (HDL) in blood and is transported to the brain across the BBB from the periphery¹³⁴. Sengupta et al. suggested ApoA-1 could bind to neurotoxic A β and reduce neuronal oxidative stress, thus playing a neuroprotective role¹³⁵. Other than these biomarkers, IGFBP-2 and fibrinogen- γ -chain were consistently upregulated in AD cohorts¹²⁷. Lastly, differential expression of pancreatic polypeptide (PP) in AD cohorts might provide insight into the relationship and involvement of PP in type 2 diabetes mellitus and the effect of abnormal insulin towards A β aggregation and neurofibrillary tangles¹³⁶. The majority of potential AD blood biomarkers identified to date are involved in inflammation, so high specificity for AD is unlikely, since many if not most diseases/disorders have an inflammatory component. Specificity of such markers needs further validation, by comparison with other diseases, neurodegenerative as well as other age related diseases (e.g., metabolic syndrome, type 2 diabetes, cardiovascular disease, etc). Neuroinflammation is a central phenomenon for several neurodegenerative diseases such as Parkinson's disease¹³⁷. There is an urgent need to conduct

more proteomics experiments to identify biomarkers with specificity to differentiate AD from other age related diseases. Considered from a somewhat different perspective, it is useful to identify processes which are shared across a variety of age-related diseases, since such processes may reflect age-related changes which predispose/contribute to disease progression. Examining the proteome as broadly as possible may also identify changes which are disease specific. With such a dual approach, biomarkers for disease risk factors (e.g., inflammation, neurodegeneration, vascular disease, metabolic syndrome, etc) might be identified and differentiated from disease specific biomarkers. Such an approach would require both a diversity of appropriate cohorts, as well as broad spectrum “omics” biomarker detection protocols.

1.6.1 Different samples used for the analysis of AD proteome:

Proteomics methods have evolved over time to examine clinical AD in order to clarify the disease's underlying biomolecular pathways. Given the difficulties associated with examining the entire proteome initially, the sub-proteome analysis looked to be an efficient technique, using a diversity of tissue types in both human and AD animal models. Table 1.3 highlights the proteomics techniques and sample types used to profile the AD proteome in previous studies. Apart from subproteome investigations, various researchers attempted to characterize the complete proteomic alterations in AD brain, CSF and blood, obtaining between <100 and 14,000 proteins, using bottom-up proteomics approaches (Table 1.3). Despite its numerous advantages, plasma/serum is one of the least investigated samples in Alzheimer's disease research, highlighting the critical necessity to examine AD plasma.

Table 1.3: This table shows the number of proteins from different sample types in AD identified by mass spectrometry-based proteomics techniques.

Sample type	Technique used	Number of proteins identified	Reference
Amyloidome from AD brain tissue	Label-free LC-MS/MS	488	Liao et al., 2004 https://pubmed.ncbi.nlm.nih.gov/15220353/
Amyloidome from AD brain tissue	Label-free LC-MS/MS	900	Drummond et al. 2017 https://pubmed.ncbi.nlm.nih.gov/28258398/
Amyloidome from AD APP/PSEN1 transgenic model mice	TMT-LC/LC-MS/MS	4,000	Xiong et al., 2019 https://pubmed.ncbi.nlm.nih.gov/30502339/
Detergent-insoluble frontal cortex AD tissue	Sequential fractionation using gel electrophoresis	512	Gozal et al. 2009 https://pubmed.ncbi.nlm.nih.gov/19746990/

	coupled with LC-MS/MS		
AD brain insoluble proteome	Label-free LC-MS/MS	4,216	Bai et al., 2013 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3799305/
Synapse AD proteome	Label-free LC-MS/MS	494	Zhou et al., 2013 https://pubmed.ncbi.nlm.nih.gov/23537733/
Synapse AD proteome	TMT-LC/LC-MS/MS	5,000	Hesse et al., 2019 https://pubmed.ncbi.nlm.nih.gov/31862015/
AD brain tissue proteome	Label-free LC-MS/MS	5,688	Johnson et al., 2020 https://pubmed.ncbi.nlm.nih.gov/32284590/
AD post-mortem brain tissue proteome	Samples were fractionated by offline basic pH reverse phase LC and analysed using TMT-LC/LC-MS/MS	14,513	Bai et al., 2020 https://pubmed.ncbi.nlm.nih.gov/31926610/
AD CSF proteome	Data-independent acquisition strategy (DIA) MS/MS	1,233	Bader et al., 2020 https://www.embopress.org/doi/full/10.15252/msb.20199356
AD plasma proteome	Olink Proteomics using PEA technology	1,160	Jiang et al., 2021 https://alzjournals.onlinelibrary.wiley.com/doi/10.1002/alz.12369
AD plasma proteome	iTRAQ/ LC-MS/MS	145	Song et al., 2014 https://proteomesci.biomedcentral.com/articles/10.1186/1477-5956-12-5
AD plasma proteome	iTRAQ/ LC-MS/MS	81	Muenchhoff et al., 2016 https://www.nature.com/articles/srep29078

1.7 Major challenges in blood-based proteomics technologies:

In the past decades, proteomics technologies have advanced to become extremely useful tools for a wide range of applications to investigate the new drug targets and early disease markers by understanding the biological systems. Mass spectrometry-based methods are more rapid, sensitive and provide greater proteome coverage than other approaches, in addition to providing an unbiased/non-selective approach, in the sense that all proteins within detection limits are identified using the non-targeted data-dependent analysis mode. The greater throughput MudPIT approaches are replacing, the slower and laborious traditional methods of top-down analysis such as 2D gels^{138,139}. However, the major challenge of investigating proteomics profiles remains the complexity of the physiological conditions and biological systems, in addition to big data processing and analysis. Several studies have investigated biofluid or tissue changes in diseased conditions; however, only a handful of proteins have been confirmed as potential targets. After years of development, current proteomics technologies are still under development and show certain limitations, including 1) the

complexity of the analysis; 2) the risk of a high rate of false positives; 3) the lack of standardisation in sample processing; 4) the dynamic range of the samples such as blood limiting the identification of low abundance proteins, and 5) the inability to validate biomarkers in a large number of patients due to a lack of antibodies.

Apart from technological advancements in mass spectrometry, sample preparation protocols can contribute greatly to level of proteome coverage. Sample fractionation is critical for complicated samples such as plasma (with a concentration range of up to 13 orders of magnitude between the lowest and highest concentration component)³⁸. Without preliminary fractionation, even the most advanced mass spectrometers are unable to achieve sufficient depth of proteome coverage on such a sample. This work can be challenging, but is necessary to achieve necessitate an increase in the sensitivity and performance of analytical techniques, particularly for the identification and quantification of specific protein classes, such as low-abundance proteins.

The abundance of a few proteins such as albumin and the immunoglobulins, representing ~75% of the total protein content in the plasma and serum, which can limit the MS detection range of medium to less abundant proteins, without prior fractionation or enrichment of lower abundance proteins¹⁴⁰. Prefractionation of serum/plasma samples is widely used to reduce sample complexity and provide better coverage of the proteome profile. Another approach is immunoaffinity-based depletion of the high abundance proteins or enrichment of the low abundance proteins to improve the detection of lower molecular weight proteins. Lack of studies on the effects of the fractionation and depletion approaches on the potential removal of low abundance proteins in addition to targeted high abundance proteins is a limitation. For example, albumin-bound lower molecular weight proteins may also be removed in addition to albumin. Another issue with profiling the plasma/serum proteome for tissue-specific proteins more likely to be clinically relevant is that they are likely to be of medium to low abundance and, therefore, more likely to be identified in fractionated plasma. It is essential to use specific prefractionation and depletion schemes and prudent adoption of methods that can both identify and quantify proteins to answer biological questions. These advances will have a significant impact on diagnostic research and development as well as pharmaceutical and biotechnology industries¹⁴¹.

Additionally, ongoing improvement of proteome throughput increases the depth of identified proteomes and enables large sample sizes to overcome constraints imposed by dynamic protein range and clinical sample heterogeneity.

1.8 Discovering early biomarkers of AD pathology:

Preclinical AD, defined as a stage of amyloid-mediated neurodegeneration occurring before the onset of clinical symptoms, is an appealing target for disease-modifying intervention in AD^{142,143}. AD development occurs during a lengthy prodromal stage prior to progressive neurodegeneration, implying the existence of a resilient mechanism against A β toxicity in the human brain, followed by amplified insults that overwhelm resilience and precipitate irreversible degeneration. By the time AD is clinically recognized, neuronal death in numerous brain and retinal locations has already happened. Thus, the availability of early and reliable disease biomarkers would enable the disease to be detected and preventive treatments be taken to avoid neuronal death. A recent study has identified some differentially expressed proteins, such as netrin-1 (NTN1), netrin-3 (NTN3), midkine (MDK), pleiotrophin (PTN), hepatocyte growth factor (HGF), and WNT5B, may serve as protective factors prior to the onset of AD, notably in human resilient instances with high A β pathology but no clinical symptoms¹⁴⁴. Due to the scarcity of research demonstrating early AD changes, there is an urgent need to incorporate longitudinal cohorts to determine which proteins are essential for illness development over the years.

1.8.1 ADAD and EOAD:

AD is typically classified into two kinds according to the age at which it manifests: early-onset AD (EOAD) and late-onset AD (LOAD), with 65 years as the generally accepted cut-off¹⁴⁵. EOAD accounts for just 5%–10% of all AD cases, with <1% caused by mutations in one of three genes: amyloid precursor protein (APP), presenilin 1 (PSEN1), or presenilin 2 (PSEN2). Generally, autosomal dominant AD (ADAD) cases indicate a "purer" form of AD with fewer coexisting diseases and thus fewer confounding variables¹⁴⁶. Whereas early-onset sporadic AD (EOsAD) is rarely investigated in isolation, it is a critical subset because it allows the discovery of novel lifestyle risk factors or the exploration of novel mutations responsible for the earlier age of onset. However, there are various reasons why biomarkers for the two disorders may differ: 1) The fact that they are pathologically similar does not necessarily suggest that the pathology is identical. 2) There are compelling reasons to believe that excessive A β ₄₂

production is an early and defining feature of ADAD. 3) The rate of development and progression of pathology is different in the two disorders such that at a particular stage of the disease (as identified by cognitive deficits), the biomarkers may differ quantitatively.

1.8.2 Risk factors for LOAD:

Late-onset AD (LOAD) is a highly complex genetic disease with a heritability between 60% and 80%¹⁴⁷. There are numerous genetic risk factors for developing LOAD, the strongest of which is the apolipoprotein E $\epsilon 4$ allele inheritance (*APOE $\epsilon 4$*)^{13,148-150}. There are three common alleles (*APOE $\epsilon 2$* , *APOE $\epsilon 3$* , and *APOE $\epsilon 4$*), resulting in six possible genotypes (*APOE 2/2*, *2/3*, *3/3*, *2/4*, *3/4*, and *4/4*). These three polymorphic alleles, i.e., $\epsilon 2$, $\epsilon 3$ and $\epsilon 4$, have a worldwide frequency of 8.4%, 77.9% and 13.7%, respectively¹³. Recent studies reported that approximately 65% of individuals with late-onset familial and sporadic AD bear the *APOE $\epsilon 4$* allele⁴. One copy of *APOE $\epsilon 4$* is associated with a threefold increase in disease risk, while two copies are associated with a more than tenfold increase in risk¹⁵. In addition, *APOE* plays a critical role in lipid transport and cholesterol homeostasis in the brain, as it does in the rest of the body^{13,151}. In the CNS, *APOE* is primarily made in astrocytes, and it facilitates the transportation of cholesterol to neurons by binding to LDLR family members, known as APOE receptors. However, approximately 25%–40% of patients with AD dementia do not carry the *APOE $\epsilon 4$* allele, and the pathophysiological mechanisms underlying AD are less clear in these individuals^{13,152}. To gain a better understanding of how *APOE* genotypes may influence AD pathology, a comprehensive proteomic analysis of human plasma samples using a series of differentially expressed proteins and molecular networks is required. Additional common genetic variations associated with late-onset AD have been found through genome-wide association studies¹⁵³. These genes may interact with one another through biological processes involving lipid metabolism, innate immunity, and endocytosis. With the advent of next-generation sequencing, it is becoming possible to identify uncommon genetic variations that have a significant effect on illness risk. For example, uncommon mutations in the triggering receptor expressed on the myeloid cells 2 (TREM2) gene are associated with an approximately three-to-fourfold increased risk of developing AD¹⁵³.

Early AD manifests clinically as mild cognitive impairment (MCI)¹⁵⁴, although a clinical diagnosis of MCI does not always progress to dementia. By the time AD manifests as dementia, the level of brain pathology is impossible to revert since substantial neuronal cell death has

occurred. Identifying biomarkers of transition from normal to MCI (if not earlier) might provide a window of opportunity for prevention trials that focus on ameliorating symptoms before neurodegeneration progresses to clinically identifiable symptoms. The transition from MCI to AD may be triggered by an increase in harmful events and a decrease in protective events, which occurs along with a significant increase in tau pathology. A recent study discovered a long list of differentially expressed proteins such as C1QL1, C1QA, C4B, C1QTNF5, C1R and others involved in the complement system and may influence the cellular and molecular events in the progression of AD¹⁵⁵. Neurotrophic factors, such as VGF, BDNF, NRN1, and CRH, were decreased during the progress from MCI to AD¹⁴⁴.

1.9 Pathways dysregulated in AD:

Major unresolved questions in AD research pertain to the interrelationships between the different pathologies. Overexpression of the mutant APP protein results in the formation of plaques but not in neuronal death in transgenic animals¹⁵⁶. These findings imply that the creation of plaques alone is insufficient to induce disease and that additional elements must be involved in human pathophysiology. Numerous mechanisms have been proposed to contribute to AD, including neuroinflammation and oxidative stress, metabolic dysfunction, reduced clearance of misfolded proteins and other less-studied mechanisms.

1.9.1 Metabolic dysfunction:

While the several genetic factors listed above contribute significantly to AD, none can explain 100% of AD cases. Therefore, rather than focusing on individual genes, it may be appropriate to explore AD due to disturbances to entire biological networks. During ageing, neurons with a low capacity for regeneration become unable to adapt to changes in the basal metabolic rate, and the energy-driven state is diminished or damaged, perhaps contributing to the development of a variety of neurodegenerative diseases. Decreased neuronal glucose metabolism and accompanying bioenergetic changes are well-known characteristics of AD¹⁵⁷. In AD, decreased glucose sensing by the brain may signal the body to enter a fasting state, triggering compensatory activation of alternative fuel sources for the TCA cycle, such as amino and fatty acids¹⁵⁸. Mitochondrial dysfunction may cause the release of reactive oxygen species, hence initiating oxidative stress, or, conversely, other oxidative stress initiators may result in mitochondrial dysfunction. Regardless of the initiator, mitochondrial insufficiency results in a

decrease in cellular ATP in CNS neurones. Increased extracellular glutamate, the primary excitatory neurotransmitter in the CNS, is another possible AD starter¹⁵⁹.

1.9.2 Clearance of misfolded proteins:

Despite the fact that all cells have multiple, well-characterized protein quality control systems to mitigate the toxicity of misfolded proteins, how they are integrated to maintain protein homeostasis ('proteostasis') in health and how their disintegration contributes to disease remains an exciting and fast-paced area of research. Autophagy is becoming recognized as a critical function in regulating neuronal and glial cell health in AD¹⁶⁰. It is a multistep process involving sequestration, degradation, and amino acid/peptide generation. It is mediated by a unique organelle called the autophagosome, a vesicle containing cellular material targeted for degradation (macromolecules and organelles) by an intracellular degradation system. Notably, there is strong evidence that autophagy is dysregulated in AD patients and animal models¹⁶¹. The researchers administered rat cortical neurons with rapamycin, lysosomal inhibitors, or vinblastine, a compound that activates autophagy and inhibits autophagic vacuoles (AVs) clearance by lysosomes. Neurons treated with lysosomal inhibitors or vinblastine accumulated AVs with similar morphologies to those found in the AD brain in the PSEN1/APP mice model. As a result, the authors concluded that autophagy dysfunction caused by defective AVs clearance, rather than autophagy induction alone, resulted in AD-like degenerative abnormalities¹⁶². In general, research continues to differ on which stage or stages of the autophagic-lysosomal pathway are defective in Alzheimer's disease.

1.9.3 Inflammation and oxidative stress:

Although acute inflammation in the brain is a well-established defence against infection, injury, and toxins, disruption of the balance of anti-inflammatory and pro-inflammatory signalling, as in AD, results in chronic inflammation¹⁶³. Numerous studies have demonstrated increased inflammatory markers in the brains of patients with Parkinson's disease (PD)¹⁶⁴, traumatic brain injury associated with chronic traumatic encephalopathy (CTE)¹⁶⁵, and amyotrophic lateral sclerosis (ALS)¹⁶⁶, to name a few key examples. Stimulation of the immune system in response to A β and proinflammatory cytokines affects the microglial clearance of A β and neuronal debris. Simultaneously, immunological activity may impair microglial production of neurotrophic factors¹⁶⁷. Collectively, phenotypic alterations in microglia contribute to cognitive impairment. It is becoming increasingly clear that persistent immunological response

is a hallmark of neurodegenerative diseases. Oxidative stress, a process that occurs more frequently in the brain as we age, is caused by a redox imbalance involving the formation of excessive reactive oxygen species (ROS) or the antioxidant system malfunctioning¹⁶⁸. Recent evidence suggests that the Hippo signalling pathway is involved in neuroinflammation, neuronal cell differentiation, and neuronal death¹⁶⁹. 14-3-3 protein family is highly expressed in the brain and influence many aspects of brain function through interactions with a diverse set of binding partners, including neural signalling, neuronal development, and neuroprotection¹⁷⁰, and is a well-studied protein family in AD CSF^{171,172}.

1.9.4 Post-translation modifications (PTMs):

The complex interplay and post-translational modifications (PTMs) of a range of proteins regulate many pathogenic processes. To discover new biomarkers for reliable diagnoses and therapeutics, it is necessary to explore the potential role of diverse PTMs in illness development. Tau is an essential regulator of the neuronal cytoskeleton and is involved in the stabilisation of microtubule assembly. The interaction of kinases and phosphatases regulates the phosphorylation status of the protein. A recent study has discovered one of the most thorough tau PTM investigations in AD brains using a variety of MS techniques¹⁷³. On numerous tau isoforms from 42 control and 49 AD patients, a total of 95 alteration events (55 in phosphorylation, 17 in ubiquitination, 19 in acetylation, and 4 in methylation) were found. Many protein phosphatases (PPs), including PP1, PP2A, PP2B, and PP5, have been demonstrated *in vitro* to dephosphorylate the tau protein at Ser199, Ser202, Thr205, Thr212, Ser214, Ser235, Ser262, Ser396, Ser404, and Ser409. Additionally, tau acetylation has been shown to increase during the early and moderate Braak phases of tauopathy, perhaps slowing tau breakdown¹⁷⁴. Acetylation of K280/K281 sites has been proposed to increase tau aggregation¹⁷⁵. Furthermore, a global investigation of protein ubiquitination in AD was published, spanning 4,291 ubiquitinated sites in 1,682 proteins, with over 800 sites changed in AD¹⁷⁶. Polyubiquitination chains (Lys11, Lys48, and Lys63) were also observed to accumulate in AD brain tissues¹⁷⁷. These comprehensive PTM datasets show the importance of PTMs modifications on the pathology of AD. Further research on PTMs modification in AD might be useful for studying biochemical signalling networks during AD development.

1.9.5 Concluding Remarks

The core AD biomarkers (A β and Tau) are included in research diagnostic criteria, and we anticipate an increase in their usage in clinical routine practice. While conventional AD diagnostic criteria are based on clinical data, additional criteria are required to detect the disease in its early stages. It is now well known that AD begins decades before clinical symptoms manifest. The ability to detect biological changes prior to clinical symptoms would enable early diagnosis and potentially alter therapy options. Recent deep proteomics studies have already profiled the brain and biofluids raising numerous novel hypotheses for a subsequent validation. However, only a small proportion of the plasma proteome has been studied to date, mostly because of the complexity of the plasma proteome and the huge quantitative dynamic range. More research is required to develop reliable blood tests that can be used in conjunction with CSF or imaging tests as non-invasive testing methods. Additionally, this chapter has identified a gap in our understanding of longitudinal alterations necessary to elucidate disease causes in AD aetiology. Understanding the basic mechanisms that occur when AD begins or worsens will aid in AD diagnosis and therapy. Therefore, in this thesis I have provided a holistic view of the AD proteomic landscape, focusing on new insights into AD pathogenesis and potential biomarkers.

2.0 Aims and Rationale

The primary objective of this thesis was to investigate plasma biomarkers that may be implicated in the pathogenesis of Alzheimer's disease (AD). To gain an in-depth understanding of plasma biomarker changes in AD, I have developed a fractionation strategy that could provide good plasma proteome coverage and identify tissue-specific proteins while being compatible with the higher sample throughput required for many clinical studies. I further applied our developed method on two different clinical cohorts to profile the plasma proteome changes in ageing, MCI and AD dementia samples.

More specific aims were:

Aim 1:

To compare the fluid biomarker patterns in early-onset subtypes of AD; early-onset sporadic AD (EOsAD) and autosomal dominant AD (ADAD). ADAD and EOsAD can be clinically and phenotypically identical, with the critical difference being the presence of an autosomal dominant mutation in the former and perhaps risk genes and environmental/lifestyle factors in the latter. Further, the focus of this chapter was also to show the relative paucity of additional biomarkers, and low numbers of reported studies in relatively accessible serum/plasma sample types were notable (chapter 2).

Aim 2:

The principal aim of chapter 3 was to identify a fractionation strategy that would provide adequate plasma proteome coverage and identify tissue-specific proteins while being compatible with the higher sample throughput required for many clinical studies. Plasma is arguably one of the most challenging sample types for identifying biomarkers in AD. The predominance of high-abundance proteins, which account for >99 percent of the total plasma protein content, might obfuscate the detection of low-abundance components by mass spectrometry. Due to the lack of robust, reproducible, and high-throughput proteomics workflows, relatively few plasma biomarkers identified by proteomics techniques have progressed to use in clinical practice. As a result, in chapter 3, I began to overcome these limitations by combining different prefractionation methods to speed the development of plasma-based biomarkers.

Aim 3:

The label-free proteomics analysis of longitudinal and cross-sectional analyses of ageing and disease-related proteomic changes in the Sydney Memory and Ageing Study (MAS) cohort. In combination with the above-stated method (chapter 3), this study design has addressed the following questions: (1) plasma profile of differentially expressed proteins in normal ageing, ageing with progression to mild cognitive decline (MCI) and AD over the time of 6 years (2) differentially expressed plasma proteome profiles of MCI and AD as compared to their age-matched normal cognitive controls which may give possible plasma biomarkers to aid in diagnosis (3) cross-sectional analysis of baseline data, when the subjects are clinically

identified as cognitively normal, provides insight into the preclinical changes which precede subsequent progression to AD (chapter 4).

Aim 4:

The main objective of this project was to understand better the changes in plasma proteome caused by *APOE* ϵ 3 and *APOE* ϵ 4 that contributes to the biomarkers profile in AD. A comprehensive proteomic analysis of human plasma samples using Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) cohort was conducted following our method optimization in chapter 3. In this study, both PiB-PET confirmed AD and control carriers of *APOE* ϵ 3 and *APOE* ϵ 4 alleles were included. Lastly, I discovered a set of potential AD dementia plasma biomarkers replicated in both cohorts, i.e., MAS and AIBL cohorts.

Chapter 2

Fluid Biomarkers and APOE Status of Early Onset Alzheimer's Disease Variants: A Systematic Review and Meta-Analysis

The majority of the content of this chapter has been published:

Gurjeet Kaur, Anne Poljak, Nady Braidy, John D Crawford, Jessica Lo, Perminder S Sachdev. Fluid Biomarkers and APOE Status of Early Onset Alzheimer's Disease Variants: A Systematic Review and Meta-Analysis. Journal of Alzheimers Disease 2020;75(3):827-843. doi: 10.3233/JAD-200052. PMID: 32333592.

See Appendix for the complete publication

2.1 Introduction

Despite intensive research over the last two decades, no reliable treatment exists to reverse or impede the onset and progression of Alzheimer's disease (AD), which is usually sub-categorized into two subtypes based on the age of onset: early-onset AD (EOAD) and late-onset AD (LOAD), with age 65 years as the generally accepted cut-off¹⁴⁵. EOAD accounts for only 5%-10% of all AD cases¹⁷⁸, with <1% being attributed to autosomal dominant AD (ADAD), caused by mutations in one of three genes: amyloid precursor protein (APP), presenilin 1 (PSEN1), and presenilin 2 (PSEN2). ADAD cases have 100% penetrance of a known genetic mutation, with predictable age of symptom onset, and therefore presymptomatic individuals can be studied several years before symptom onset. ADAD cases generally represent a "purer" version of AD with fewer concomitant pathologies and hence confounding variables¹⁴⁶. Early-onset sporadic AD (EOsAD) is rarely studied in its own right and is, therefore, a particularly important subset as it presents an opportunity to either discover new lifestyle risk factors or explore novel mutations responsible for the earlier age of onset. Genetic variations, particularly that of the apolipoprotein E (*APOE*) gene, play a significant role in LOAD but are poorly understood in EOsAD. ADAD and EOsAD can be similar pathologically and phenotypically, with the primary distinction being the presence of the autosomal dominant mutation in the former and possibly risk genes and environmental/lifestyle factors in the latter.

However, there are several reasons why biomarkers may differ for the two disorders: 1) That they are pathologically similar does not imply that the pathology is exactly the same. Considering the difference in aetiology, it is possible that there are subtle differences in the pathology, e.g. the relative amyloid and tau burden, the degree of inflammation, oxidative stress and synaptic loss may differ, which may be reflected in quantitative differences in biomarkers; 2) There are good reasons to believe that over-production of A β ₄₂ is an earlier and defining feature of ADAD pathophysiology, but one could argue that this may not be the central disturbance in EOsAD, which would be reflected in the biomarkers, in particular, A β levels; and 3) The rate of development and

progression of pathology is different in the two disorders such that at a particular stage of the disease (as identified by cognitive deficits), the biomarkers may differ quantitatively.

Therefore, I started with the hypothesis that the biomarkers would be different. However, had I set up a null hypothesis, i.e., there would be no difference in the biomarkers for the EOAD and ADAD at similar levels of cognitive deficit, our methodology or conclusions would not have been different. The failure of numerous clinical trials suggests that a focus on presymptomatic biomarkers with a view to prevention may be a better option than treatment once pathology has significantly progressed.

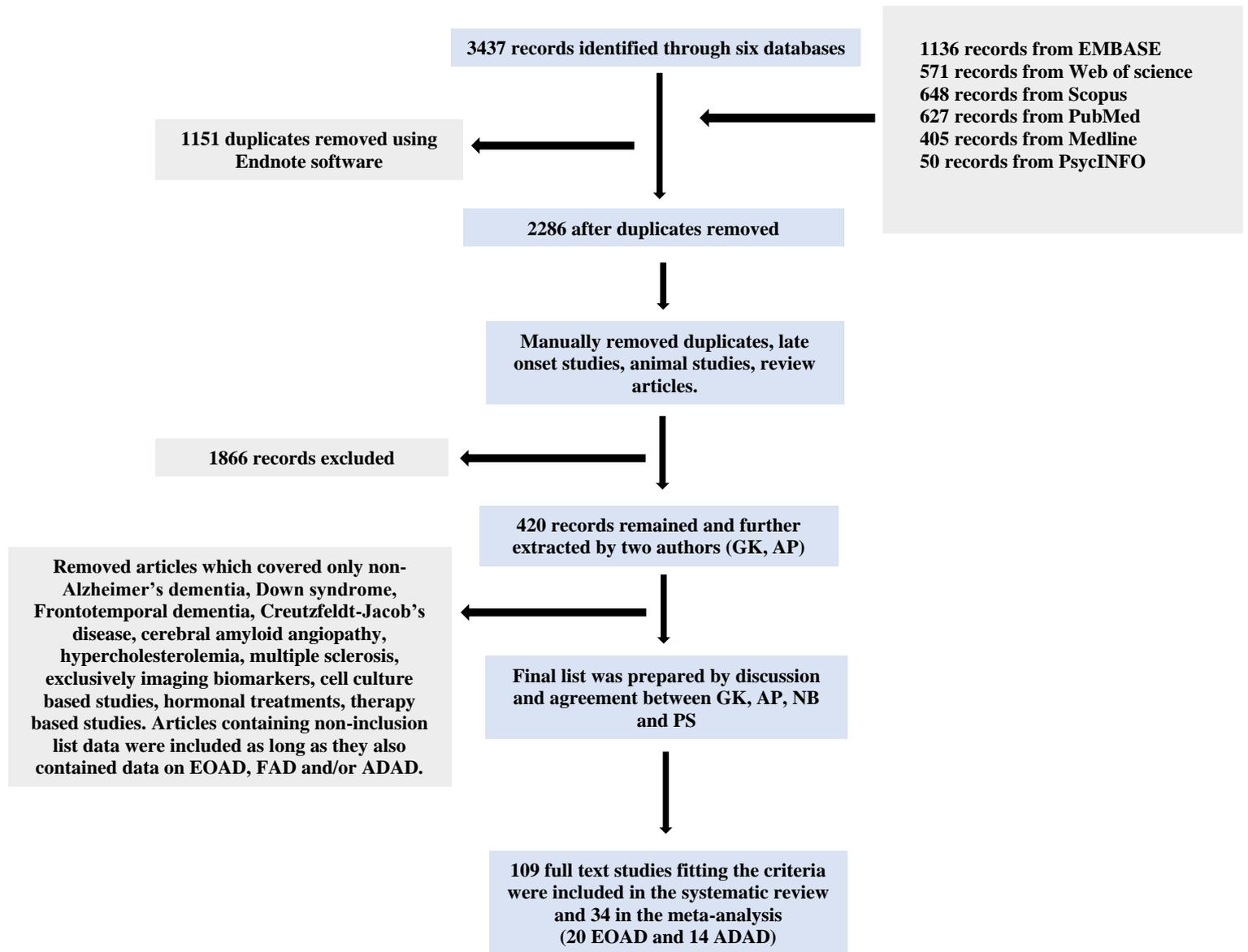
The absence of comparative studies reporting pathological and phenotypical difference/s in ADAD and EOAD makes it challenging to know whether there are qualitatively or quantitatively similar or different biomarker patterns in these genetically different AD subtypes. Therefore, I chose to perform a systematic review and meta-analysis on variants of EOAD to examine the shared and unique factors that influence subtypes of EOAD.

2.2 Methods

2.2.1 Search strategy

This systematic review and meta-analysis followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines¹⁷⁹ and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines⁹. Literature searches were performed for papers published between January 1, 1988, and June 1, 2019, using six major scientific databases: Medline Ovid, PubMed, EMBASE, PsycINFO, Web of Science and Scopus (Figure 2.1).

Figure 2.1: Flow chart depicting the selection process of records found using six scientific databases and elimination of non-relevant articles.



2.2.2 Data extraction

Inclusion/Exclusion criteria

Included in the review were papers reporting on quantified biomarkers (proteins/peptides) in CSF and/or blood (serum or plasma) in patients with EOAD or ADAD and corresponding controls. These searches included combinations of the following Boolean terms: [blood OR serum OR plasma OR cerebrospinal fluid OR CSF] AND [autosomal dominant OR dominant inheritance OR early-onset OR young-onset Alzheimer's disease] and were run in all six

databases. Exclusion criteria were: *(a)* studies purely on LOAD or on any AD subtype with biomarker data on tissues other than blood, serum, plasma or CSF, studies reporting purely on non-protein biomarkers such as mRNA (Q-PCR) or studies that use non-quantitative observations (e.g., SDS-PAGE without identifying or quantifying specific protein changes); *(b)* studies purely on animal models, cell culture techniques, immunohistochemistry, genotyping, imaging (PET, FDG), laboratory tests such as blood cell counts, and identification of biomarkers from cellular blood fractions such as peripheral blood cells; *(c)* articles reporting data following treatment, therapy, intervention, toxicology or human *in vivo* challenge studies, with either endogenous or synthetic/pharmaceutical compounds and without baseline data, and *(d)* studies on diseases with early-onset dementia as a component of the phenotype but of predominantly different aetiology (e.g., vascular). Given the limited volume of literature on this topic, no studies were excluded based on cohort size and case studies were included in the systematic review, though not in the meta-analysis. However, only cross-sectional studies were available in sufficient numbers to allow meta-analysis.

The meta-analysis included 34 studies, 20 for EOsAD and 14 for ADAD (which included both symptomatic and asymptomatic mutation carriers, sMC and aMC, respectively). By definition, the EOsAD individuals received a clinical diagnosis of AD with onset before 65 years, and the majority of them had no family history of EOAD. Genetic testing was generally not performed to eliminate APP, PSEN1 and PSEN2 mutations; however, in 4 of the 20 EOsAD studies, genetic testing was indeed carried out, and meta-analyses were also performed on these separately. The ADAD individuals had a similar age of onset, a positive family history of ADAD, and carried pathogenic mutation/s to one of the following genes: APP, PSEN1 and PSEN2. Across all studies, biomarker level changes were expressed as standardized mean differences (SMD), calculated relative to levels in cognitively healthy age-matched controls.

Five articles identified as meeting inclusion criteria were published on participants from the Dominantly Inherited Alzheimer's Network (DIAN) cohort. After discussion with DIAN authors, and in order to avoid duplication of data on the same participants, I included only the most recently published DIAN article¹⁸⁰, excluding the previous four from the meta-analysis where the biomarker data was duplicated^{11,12,181,182}. For articles with insufficient data, or data presented in an inaccessible manner, I emailed authors requesting additional information. The majority either did not respond or no longer had access to data. One author provided data which is included in our analyses⁹⁵.

While I did not use assay type as a selection criterion (either inclusion or exclusion) for the studies reported here, the majority (>95%) of studies used commercial kits which had been manufacturer validated. Of the 27 studies which reported assays of A β and/or tau variants, 26 used the Innogenetics (Fujireibo-Europe) xMAP technology-based ELISA kits. This fortuitous “standardisation” is likely to minimise inter-laboratory variation for the meta-analysis results of A β and tau variants reported here.

2.2.3 Meta-analyses

Meta-analyses were conducted using the procedure *metan*, within the statistical program Stata, version 15.0, with the random-effects model and the inverse-variance weighting of studies. The type-one error rate was set at 0.05. Meta-analyses were performed if there were at least two independent published studies on a biomarker, but in most cases included 3-13 studies.

For studies with continuous biomarker level as the outcome variable, data input into the statistical program were the number of cases, means and standard deviations (SDs) of the biomarker for each AD group and their corresponding control groups. The output of this program expresses the levels of biomarkers as the standardized mean differences (SMDs), equal to the difference in concentrations between the particular AD group and its control group, divided by the pooled SDs.

For studies with a binary outcome measure (such as $\epsilon 4$ carrier status), data input were the natural logarithms of the odds ratios (log ORs) and their standard errors. These were calculated as follows. If a and b are the number of $\epsilon 4$ carriers in the AD and control groups, respectively, and c and d are the corresponding numbers of non-carriers in those groups, then the odds ratio is given by $OR = a*d/b*c$. The standard errors of each log OR are calculated as the square root of $(1/a + 1/b + 1/c + 1/d)$.

The random-effects model was used for all of the meta-analyses based on the assumption that each study produces estimates of the true effects for that particular study population, but that the study populations vary across individual studies due to differences in populations, study design, and assay platforms¹⁸³. The pooled effect sizes produced by meta-analyses using the random-effects model are the estimates of the mean effect sizes for all potential studies

satisfying the selection criteria of the review. Publication bias was assessed using the Stata command *metabias*, which was employed to produce funnel plots and to perform the Egger's test to check for bias for meta-analyses comprising ≥ 8 studies. For meta-analyses comprising fewer individual studies, it was impossible to draw reliable conclusions from visual inspection of the funnel plots or Egger's test¹⁸⁴. Comparison of pooled effects from pairs of meta-analyses was achieved using the Stata procedure, *metareg*.

Two articles^{185,186} presented data as median and range, and in these cases, I substituted median for mean and derived SD from range using the following approach: Range/4 (if $n < 70$) or Range/6 (if $n > 70$)¹⁸⁷. For two articles, mean and SD values were estimated from bar graphs^{181,188}. Two articles reported serum NfL in ADAD individuals^{189,190}, although, for one of them¹⁸⁹, I could not get access to the data and could therefore not perform a meta-analysis on serum NfL. A significant share of the ADAD data came from two large cohorts: the DIAN cohort and the Alzheimer's Prevention Initiative (API) cohort representing a large pedigree living in the state of Antioquia in Colombia, South America. The DIAN cohort includes carrier and non-carrier (NC) family members with various ADAD mutations, while the Colombian kindred likely descended from a single individual and carried the E280A mutation in the PSEN1 gene.

2.2.4 Data availability statement

Data are available to qualified investigators on request to the corresponding author. Data will be shared at the request of other investigators for purposes of replicating procedures and results.

2.3 Results

2.3.1 Protein/peptide biomarkers

2.3.1.1 Early-onset Alzheimer's Disease (EOAD)

I identified 34 eligible published studies on EOAD, comprising 20 studies on EOAD and 14 on ADAD. The results for established biomarkers ($A\beta_{42}$, T-tau and P-tau) were as follows: All variants of EOAD had lower CSF $A\beta_{42}$ and higher CSF T-tau and P-tau (Table 2.1, Figures 2.2, 2.3 and 2.4), relative to their respective comparison groups. The majority of EOAD

biomarkers were identified in CSF, and few studies of blood biomarkers were replicated frequently enough to allow meta-analysis.

Table 2.1: A) Pooled standardized mean difference (pooled SMD), I-squared (I²), and significance (p) are shown with p≤0.05 accepted as statistically significant. B) Meta-analysis of APOE ε4 carriers in early-onset AD pooled odds ratio (pooled OR), I-squared (I²), and significance (p) are shown with p≤0.05 accepted as statistically significant (random-effects model is represented in the tables). Meta-analysis of CSF and serum/plasma biomarkers of EOsAD and ADAD.

(A)	Biomarker	No. of independent studies	EOAD (total numbers)	Controls (total numbers)	Meta-Analysis Random effects model
(i): Early-onset sporadic AD (EOsAD)	CSF Aβ ₄₂	8	266	211	Pooled SMD = -1.83, p < 0.001, I ² =80.9%
	CSF T-tau	10	362	247	Pooled SMD = 1.24, p < 0.001, I ² =80.6%
	CSF P-tau	13	375	321	Pooled SMD = 1.39, p < 0.001, I ² =90%
	CSF P-tau ₁₈₁ & 199	4	128	142	Pooled SMD = 1.17, p = 0.018, I ² =92.2%
	CSF P-tau ₁₈₁	3	95	102	Pooled SMD = 0.80, p = 0.11, I ² =90.5%
	CSF NFL	2	72	59	Pooled SMD = 1.42, p < 0.001, I ² =0.0%
	CSF IL6	2	17	39	Pooled SMD = -3.56, p < 0.385, I ² =95.2%
	CSF IgG	4	63	62	Pooled SMD = -0.33, p = 0.075, I ² =0.0%
	CSF albumin	3	52	39	Pooled SMD = -0.94, p = 0.241, I ² =90.9%
	Serum/Plasma albumin	3	52	39	Pooled SMD = 1.66, p = 0.351, I ² =96.8%
Serum/Plasma IgG	4	63	62	Pooled SMD = 1.8, p = 0.037, I ² =93.1%	
(ii): EOsAD with absence of AD mutations confirmed	CSF Aβ ₄₂	4	136	112	Pooled SMD = -2.18, p < 0.001, I ² =0.0%
	CSF T-tau	3	103	91	Pooled SMD = 1.66, p < 0.001, I ² =85.5%
	CSF P-tau	4	136	117	Pooled SMD = 1.45, p < 0.001, I ² =77.7%
(iii): EOsAD with no genetic testing reported	CSF Aβ ₄₂	4	130	99	Pooled SMD = -1.45, p < 0.001, I ² =87.5%
	CSF T-tau	7	259	156	Pooled SMD = 1.05, p < 0.001,

					$I^2=78.1\%$
	CSF P-tau	8	206	164	Pooled SMD = 1.24, p < 0.001 , $I^2=91.7\%$
(iv): ADAD All MCs	CSF Aβ₄₂	12	480	441	Pooled SMD = -2.11, p < 0.001 , $I^2=94.0\%$
sMCs	CSF Aβ₄₂	10	303	344	Pooled SMD = -2.36, p < 0.001 , $I^2=95.3\%$
aMCs	CSF Aβ₄₂	3	81	90	Pooled SMD = -1.64, p = 0.055 , $I^2=91.8\%$
All MCs	CSF Aβ₄₀	3	81	120	Pooled SMD = -0.81, p = 0.044 , $I^2=79.4\%$
All MCs	CSF T-tau	10	318	288	Pooled SMD = 1.53, p < 0.001 , $I^2=83.5\%$
sMCs	CSF T-tau	3	16	44	Pooled SMD = 2.97, p < 0.001 , $I^2=23.5\%$
aMCs	CSF T-tau	5	41	55	Pooled SMD = 1.67, p = 0.003 , $I^2=75.9\%$
All MCs	CSF P-tau	5	283	267	Pooled SMD = 1.63, p < 0.001 , $I^2=75.6\%$
All MCs	Plasma Aβ₄₂	2	164	115	Pooled SMD = 0.69, p < 0.001 , $I^2=52.5\%$
(B)	Biomarker	No. of independent studies	Patients APOEϵ4+ Patients APOEϵ4- (%enrichment)*	Controls APOEϵ4+ Controls APOEϵ4- (%enrichment)*	Odds Ratio
EOsAD	APOEϵ4	4	60 81 (43%)	35 83 (30%)	Pooled OR = 1.72 p = 0.044 , $I^2=0.0\%$
ADAD	APOEϵ4	3	17 43 (28%)	25 62 (29%)	Pooled OR = 1.08, p = 0.838 , $I^2=0.0\%$

*LOAD meta-analyses report APOE ϵ 4 allele enrichment in the 13.7% - 40% range.

(i) Early-onset sporadic AD (EOsAD); defined as the age of onset < 65 years and no family history of AD.

(ii) EOsAD with the absence of APP, PSEN1 or PSEN2 mutations confirmed by genetic testing

(iii) EOsAD with no genetic testing reported (diagnosis based only on the absence of family history)

(iv) Autosomal dominant AD (ADAD); EOAD with the carrier of at least one APP, PSEN1 or PSEN2 mutations confirmed by genetic testing.

In addition to established biomarkers, quantitative data on several biomarkers of neurodegeneration and inflammation were reported in 2 to 4 studies each (Table 2.1). These were CSF NfL, IL-6, IgG, albumin, serum albumin and IgG (Table 2.1 and Figure 2.2). Of these non-classical biomarkers (i.e., protein/peptide biomarkers other than A β ₄₂, T-tau and P-tau), the only statistically significant results were for CSF NfL (pooled SMD = 1.42) and serum/plasma IgG (pooled SMD = 1.80) in the subset of all those with EOAD.

Most of the meta-analyses had high levels of heterogeneity ($I^2 > 75\%$), reflecting a large amount of scatter around the pooled SMD axis. However, the direction of change, if not its degree, was consistent in >80% of established biomarker studies. Methodological differences across studies likely accounted for this heterogeneity, and therefore the random-effects model assumptions were the most appropriate¹⁸³.

Since this model permits small studies to contribute to outcomes, I explored such publication bias using funnel plots and Egger's test. Funnel plots of 5 biomarkers were performed as they had 8 or more individual studies each (Figure 2.5). Although visual inspection of the funnel plots did not give the impression of perfect symmetry around the vertical axis, the results from Egger's test implied that there was no statistically significant systematic relationship between the results of each study and its size.

Comparisons of meta-analysis results of established CSF biomarkers were performed between pairs of patient groups vs controls to examine whether classical biomarker profiles of EOAD and subsets of EOAD are significantly different to those of ADAD. All the results of the meta-regression analyses were not significant, except for T-tau in EOAD vs sMCs of ADAD, suggesting that the classical biomarker profile of EOAD is not significantly different to the ADAD even though the majority of EOAD individuals have no family history of early-onset AD, and some articles on EOAD have even eliminated the possibility of APP, PSEN1 or PSEN2 spontaneous mutation/s by genetic testing.

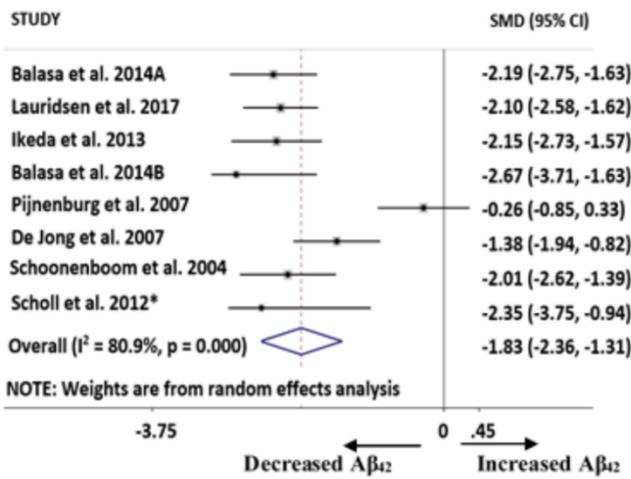
2.2.3.2 Early-onset sporadic Alzheimer's disease (EOsAD)

The 20 EOsAD studies collectively included 11 biomarkers (Table 2.1i and Figure 2.2) which had been assayed in at least two independent studies (majority >3) and on which meta-analyses could be performed. Significantly lower CSF levels of A β ₄₂ and higher levels of T-tau, P-tau and specific variants of P-tau: P-tau₁₈₁ and ₁₉₉ were observed in EOsAD vs control (Table 2.1i, Figure 2.2). Additional CSF biomarkers, with sufficient numbers of articles (≥ 2 articles) to allow meta-analysis, included NfL, IgG, IL6 and albumin (Table 2.1i, Figure 2.2), of which only NfL had a significantly higher level, while the others had no significant difference relative to controls. In serum, only albumin and IgG were assayed in sufficient numbers of studies to allow meta-analysis, and only IgG had a statistically significant result (Table 2.1i, Figure 2.2E), is increased in AD relative to controls.

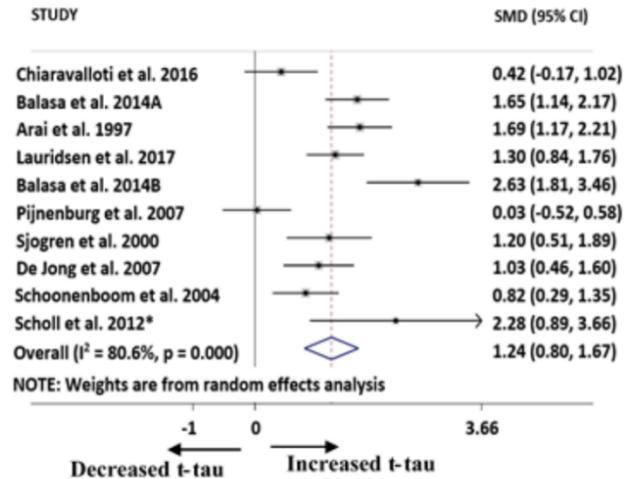
Of the 20 studies on EOsAD, a subset of 4 was identified in which APP, PSEN1, or PSEN2 gene mutations had definitively been ruled out by genetic testing (Table 2.1ii, Figure 2.3A, 2.3B, 2.3C). A separate subset of 16 studies did not report confirmation of the absence of these mutations by genetic testing (Table 2.1iii Figure 2.3D, 2.3E, 2.3F). Meta-analyses were performed separately on these two sets of studies. As shown in Table 2.1, pooled SMD values for A β ₄₂ were lower in the group in which the absence of mutations had been confirmed by genetic testing, while those for both T-tau and P-tau were higher in this group. However, none of these differences was statistically significant when examined using the Stata procedure, *metareg*.

Figure 2.2: Forest plots of early onset sporadic Alzheimer's disease patients vs controls; (A) CSF A β ₄₂, SMD $p < 0.001$, (B) CSF t-tau, SMD $p < 0.001$ (C) CSF p-tau, SMD $p < 0.001$, (D) CSF NfL, SMD $p < 0.001$, and (E) serum/plasma IgG, SMD $p = 0.037$. (F) CSF P-tau₁₈₁&₁₉₉, SMD $p = 0.018$, (G) CSF P-tau₁₈₁ SMD $p = 0.011$, (H) CSF albumin, SMD $p = 0.241$, (I) CSF IL-6, SMD $p = 0.385$, and (J) CSF IgG, SMD $p = 0.075$, (K) serum albumin, SMD $p = 0.351$. * values estimated from bar graph. p value represents the statistical significance of SDM test. References for the manuscripts represented in these forest plots are shown in the supplementary section where the full list of manuscripts used for meta-analysis appear in supplementary tables S2 and S3. Ikeda et al 2013 appears twice in panel C and F (Ikeda et al 2013 and Ikeda et al 2013*) as this study analyzed two different isoforms of P-tau i.e. P-tau₁₈₁ and P-tau₁₉₉ respectively.

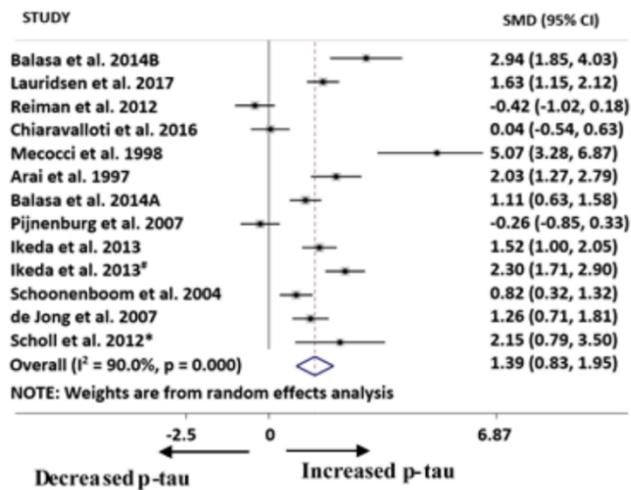
(A)



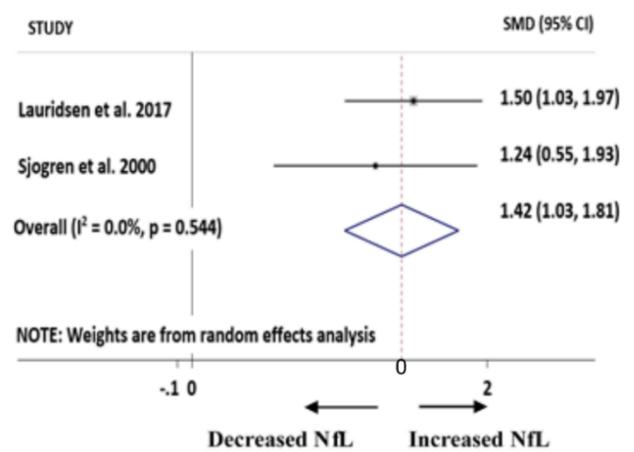
(B)



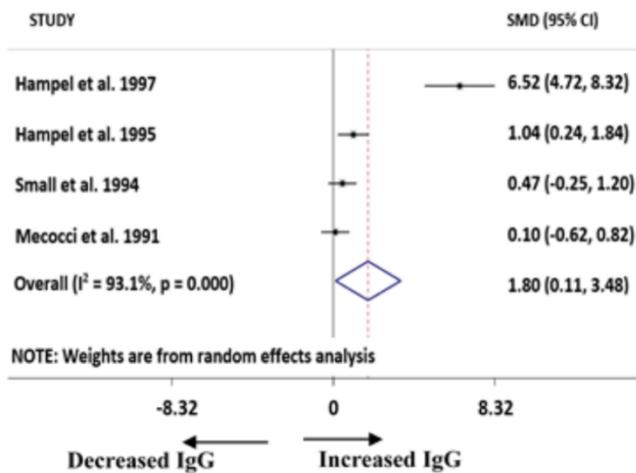
(C)



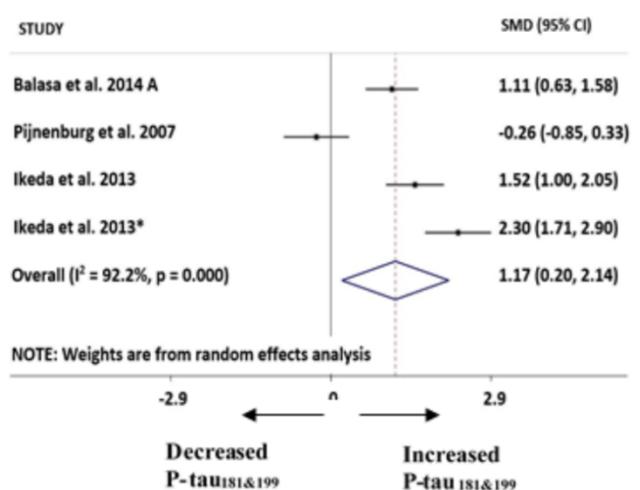
(D)



(E)



(F)



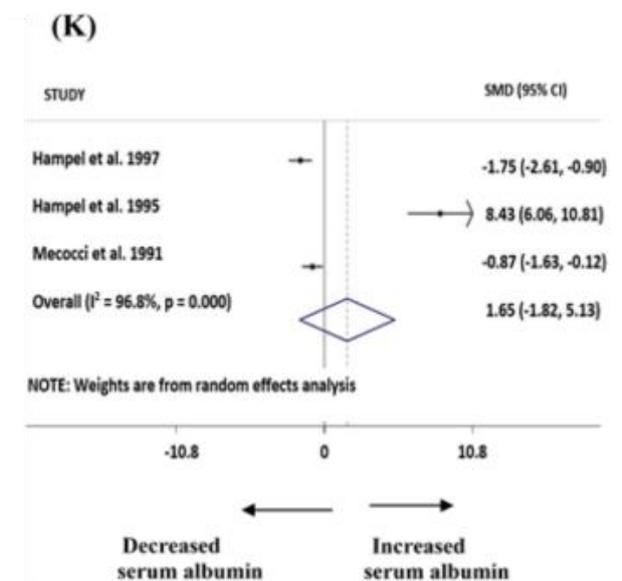
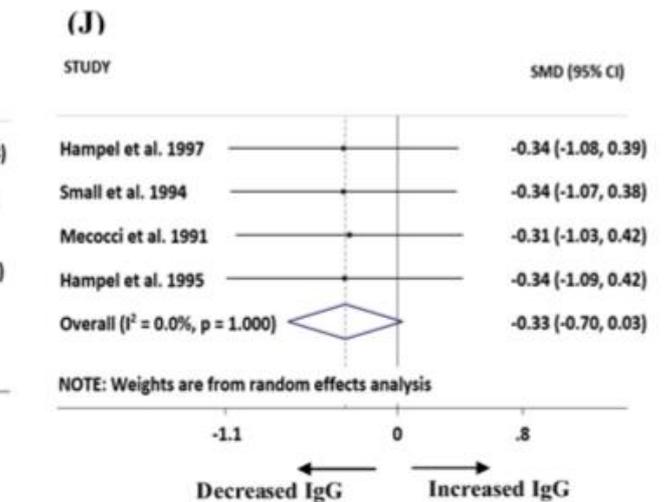
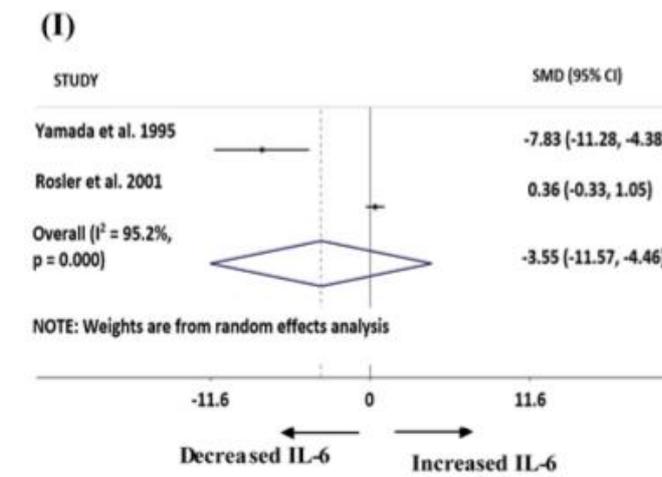
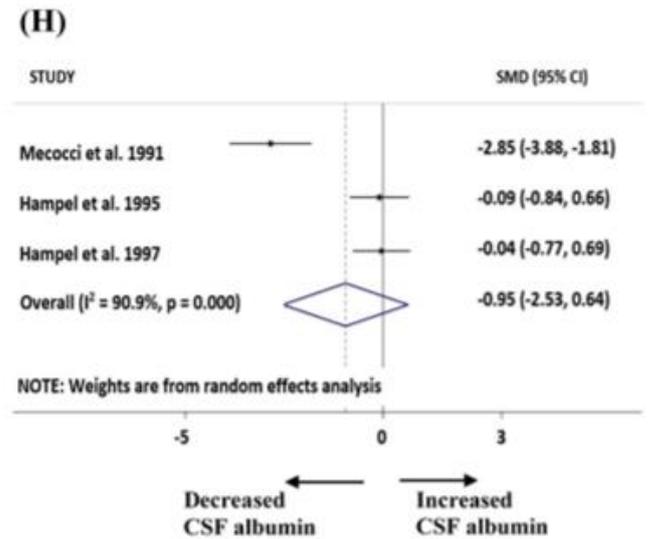
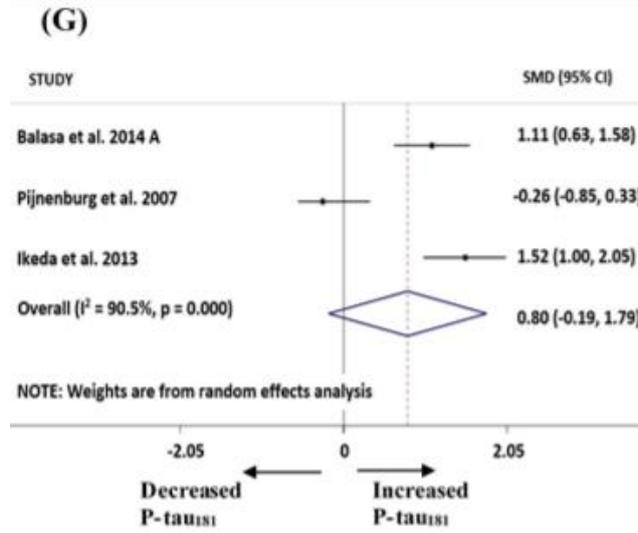
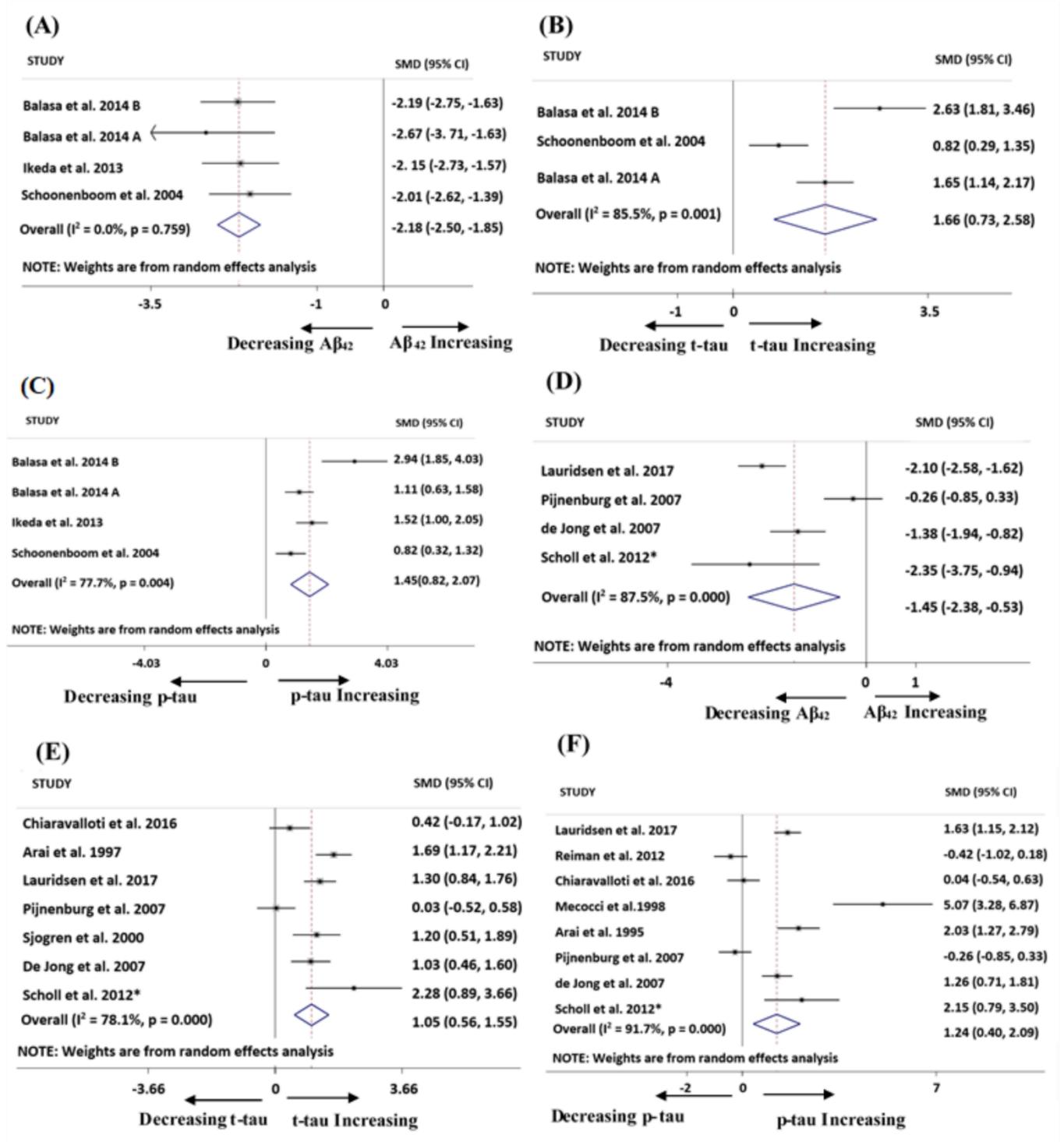


Figure 2.3: Forest plots of early onset sporadic Alzheimer’s disease patients (tested negative for APP/PSEN mutation carriers) vs normal control individuals; (A) CSF Aβ₄₂ (pg/ml), (B) CSF t-tau (pg/ml) (C) CSF p-tau (pg/ml) and early onset sporadic Alzheimer’s disease patients (no APP/PSEN mutation test has been performed) vs normal control individuals; (D) CSF Aβ₄₂ (pg/ml), (E) CSF t-tau (pg/ml), (F) CSF p-tau (pg/ml)

**values estimated from bar graph.*

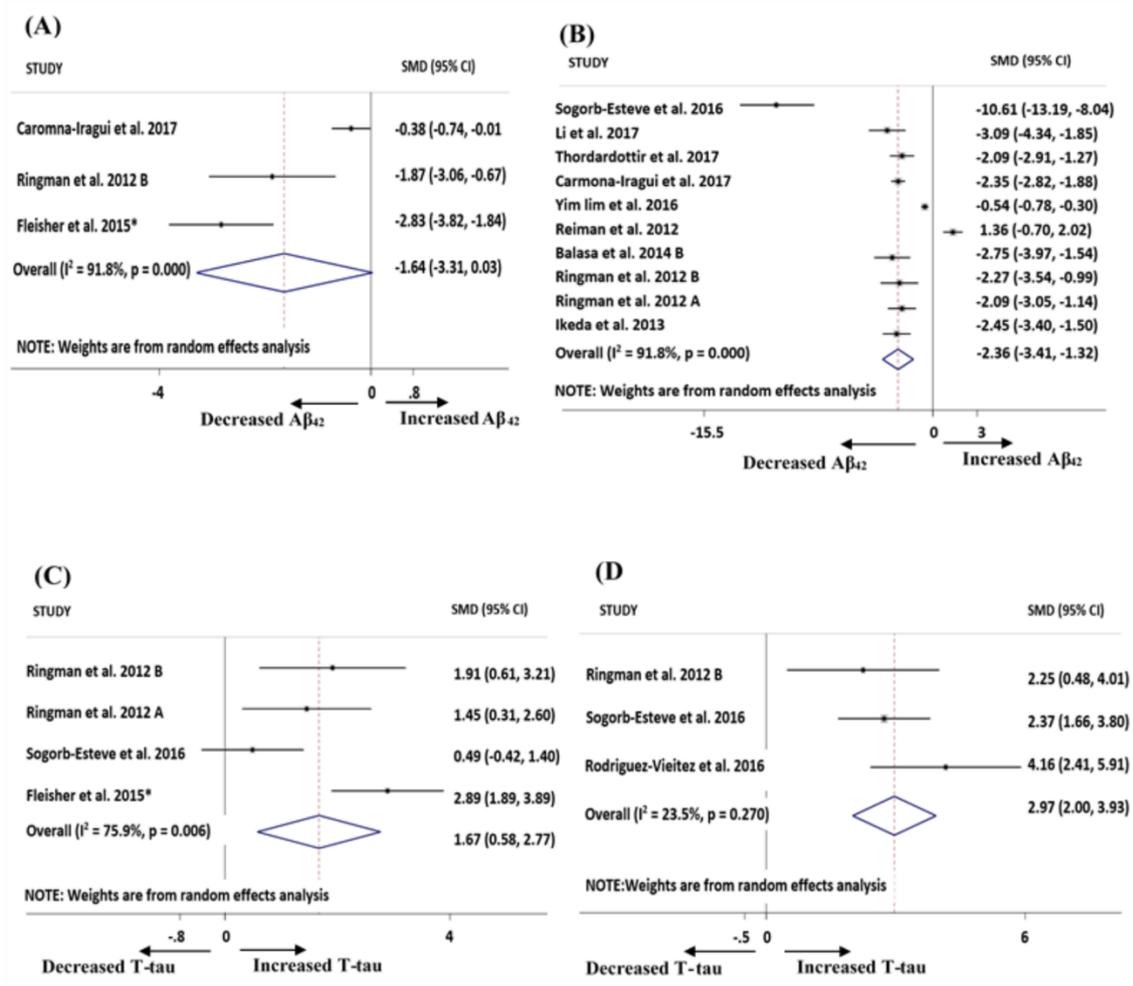


2.2.3.3 Autosomal dominant Alzheimer's disease (ADAD)

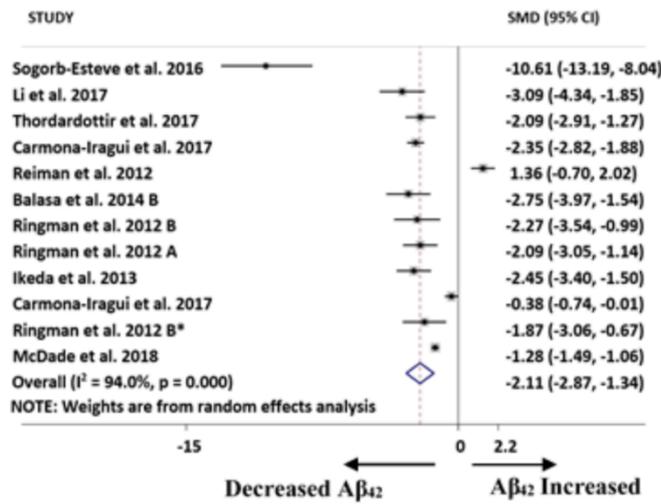
Individuals with a family history of AD and who also carry an AD-related mutation (APP/PSEN) are by definition autosomal dominant and may be identified at the asymptomatic mutation carrier (aMCs) stage or the symptomatic mutation carrier (sMCs) stage. To maximize statistical power, I initially performed a meta-analysis on all mutation carriers (all MCs), regardless of stage and subsequently stratified by symptom status (aMCs or sMCs) (Table 2.1(iv) and Figure 2.4). AD vs control comparisons for the established biomarkers was statistically significant in all MCs, with lower CSF A β ₄₂ and A β ₄₀, higher CSF T-tau, and P-tau and higher plasma A β ₄₂ for AD (Table 2.1, and Figure 2.4). Following stratification by symptom status, CSF A β ₄₂ and T-tau were significantly lower and higher, respectively, in sMCs vs controls (Table 2.1iv, Figure 2.4B, 2.4D). In aMCs, CSF A β ₄₂ also trended downwards, but pooled SMD was not as low as in symptomatic carriers and did not reach statistical significance (Table 2.1iv, Figure 2.4A). The level of T-tau significantly increased in aMCs; however, pooled SMD was not as high as in sMCs, and results were not statistically significant (Table 2.1iv, Figure 2.4C). No replicate studies of other biomarkers of AD pathology were reported in ADAD cohorts.

Fagan et al. 2014 appear twice in the panel I (Fagan et al. 2014 and Fagan et al. 2014^{*}) as this study analyzed A β ₄₂ in asymptomatic mutation carriers and asymptomatic mutation carriers, respectively. Ikeda et al. 2013 appear twice in panel G (Ikeda et al. 2013 and Ikeda et al. 2013^{*})¹⁹¹ as this study analyzed two different isoforms of P-tau, i.e. P-tau₁₈₁ and P-tau₁₉₉, respectively.

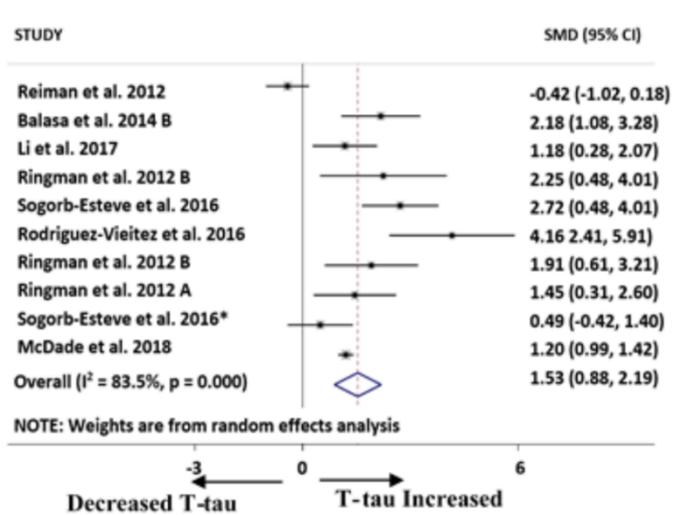
Figure 2.4: Forest plots of autosomal dominant Alzheimer’s disease patients vs normal control individuals; (A) CSF $A\beta_{42}$ asymptomatic mutation carriers, $SMD p < 0.001$, (B) CSF $A\beta_{42}$ symptomatic mutation carriers, $SMD p < 0.001$, (C) CSF t -tau asymptomatic mutation carriers, $SMD p = 0.003$, (D) CSF t -tau symptomatic mutation carriers, $SMD p < 0.001$, (E) CSF $A\beta_{42}$ all mutation carriers, $SMD p < 0.001$, (F) CSF T -tau all mutation carriers, $SMD p < 0.001$, (G) CSF P -tau all mutation carriers, $SMD p < 0.001$, (H) CSF $A\beta_{40}$ all mutation carriers, $SMD p < 0.044$, (I) Plasma $A\beta_{42}$ all mutation carriers, $SMD p < 0.001$, $APOE\epsilon 4$ allele enrichment in (J) early onset sporadic Alzheimer’s disease, $SMD p = 0.044$, (K) autosomal dominant Alzheimer’s disease patients relative to their normal controls $SMD p = 0.838$. *values estimated from bar graph.



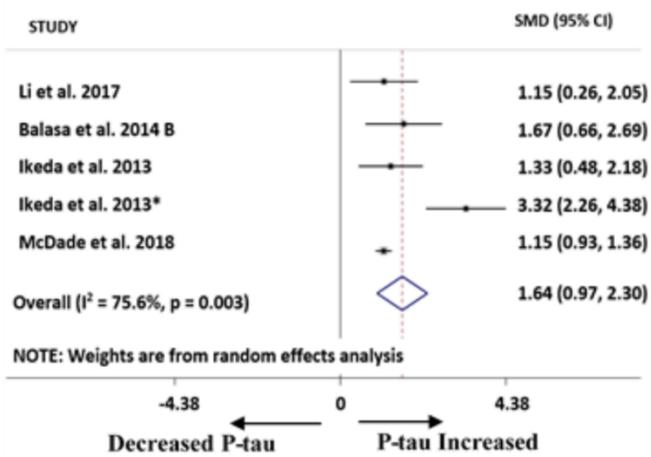
(E)



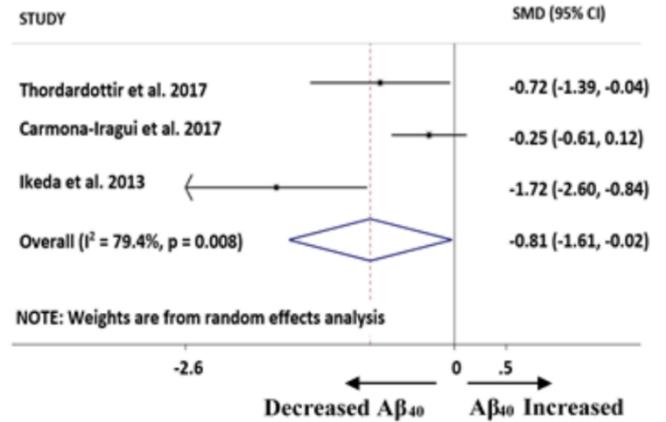
(F)



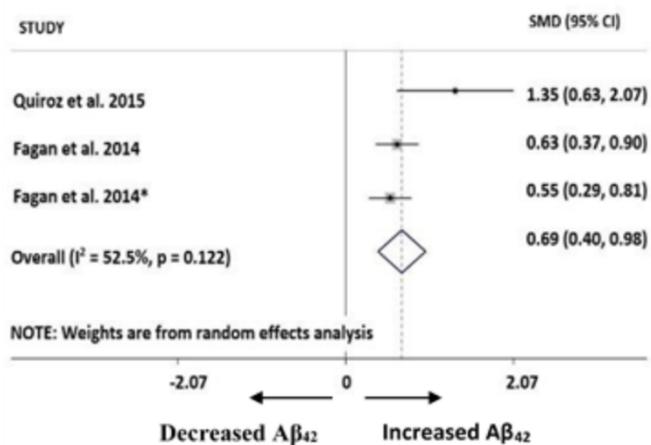
(G)



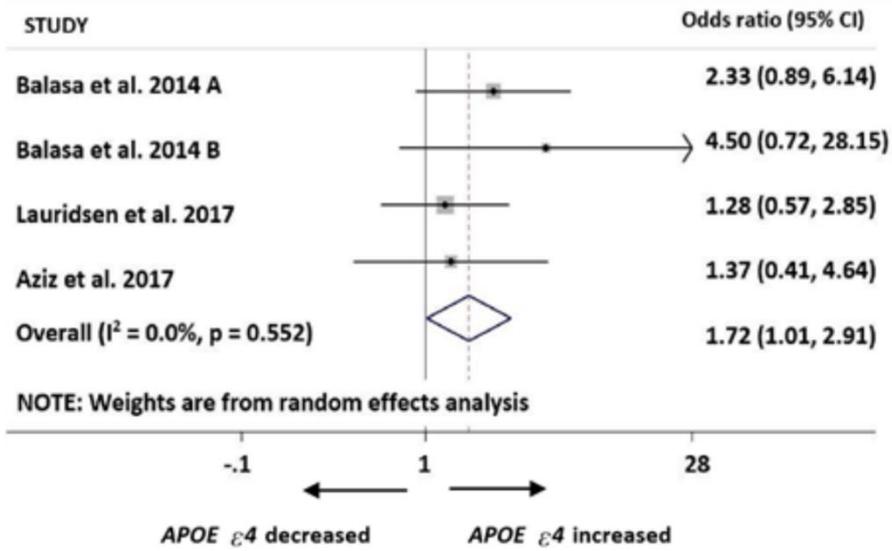
(H)



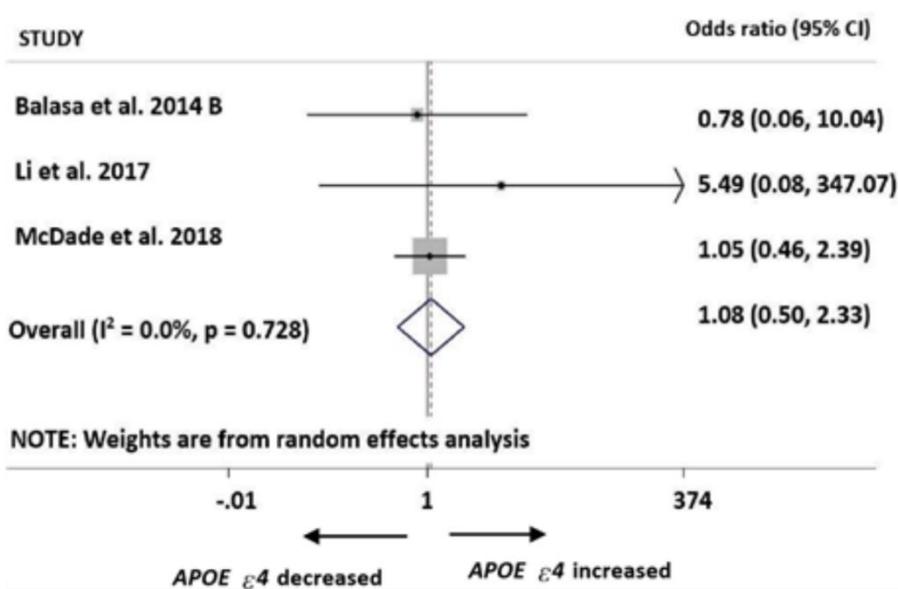
(I)

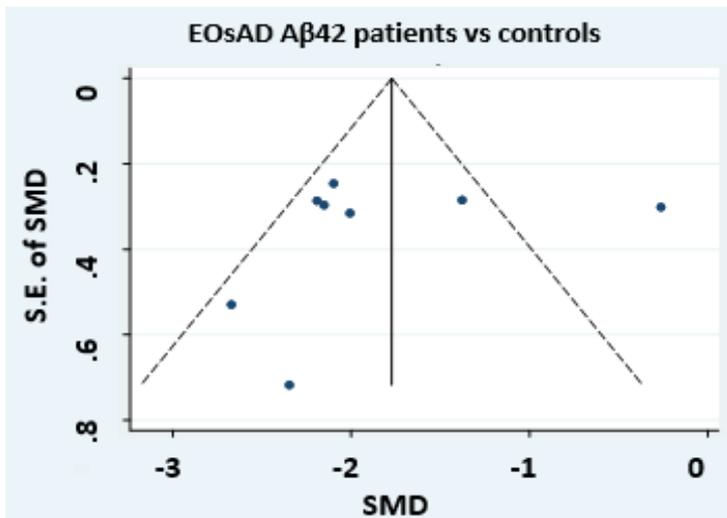


(J)

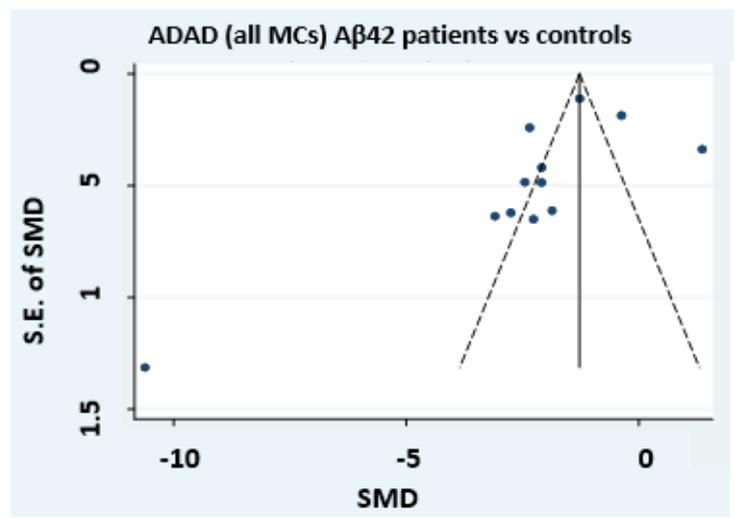


(K)

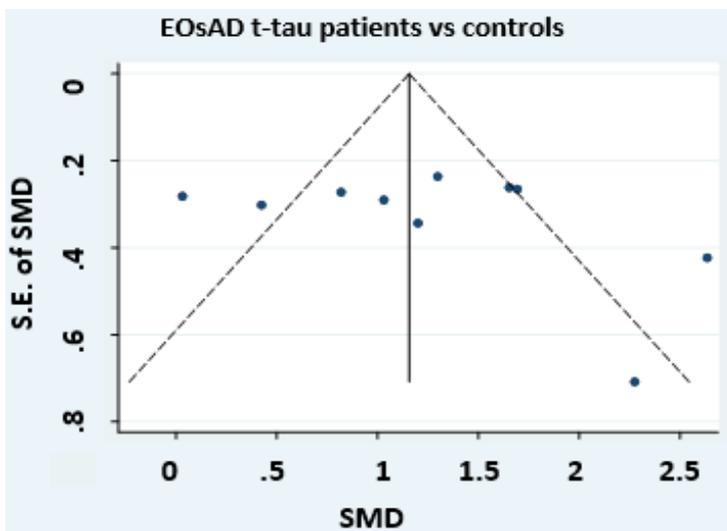




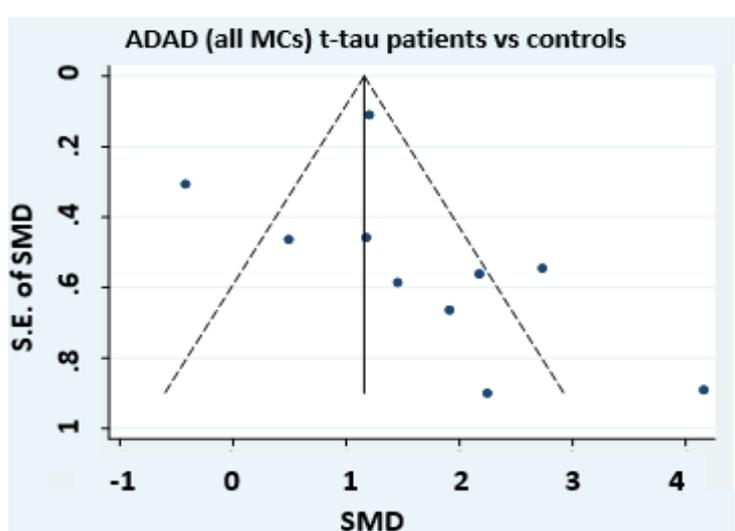
Egger's test: $P=0.64$



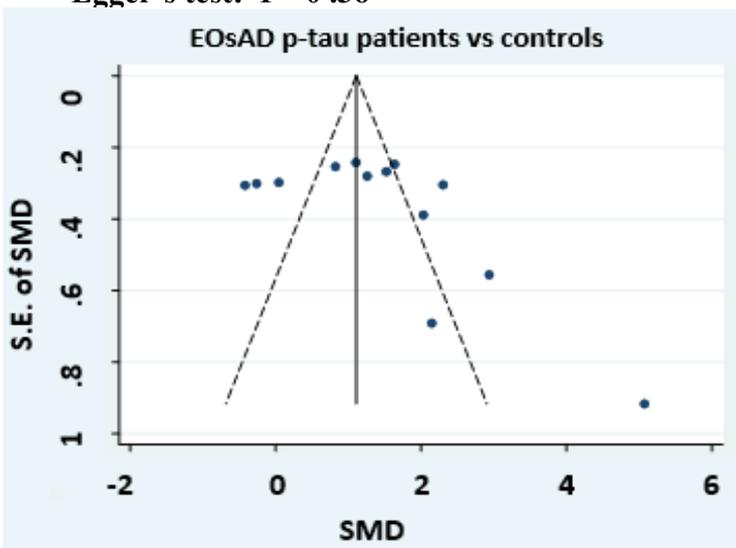
Egger's test: $P=0.15$



Egger's test: $P=0.36$



Egger's test: $P=0.39$



Egger's test: $P=0.13$

Figure 2.5. Funnel plots and Egger's test for publication bias were performed on all meta-analyses, including 8-13 individual studies. For meta-analyses comprising fewer individual studies, it is harder to make conclusions about distribution around the SDM axis, and the p-value from Egger's test is unreliable, so funnel plots are not included for meta-analyses with lower study numbers

APOE genotype status

2.2.3.4 APOE ϵ 4 allele enrichment in EOAD groups

A significantly higher proportion of EOAD participants were carriers of the *APOE ϵ 4* allele relative to their controls (Table 2.1B and Figure 2.4J). There was no significant difference in *APOE ϵ 4* prevalence between ADAD and controls (Table 2.1B and Figure 2.4K). The overall presentation of complete meta-analysis is provided in the illustration shown below (Figure 2.6).

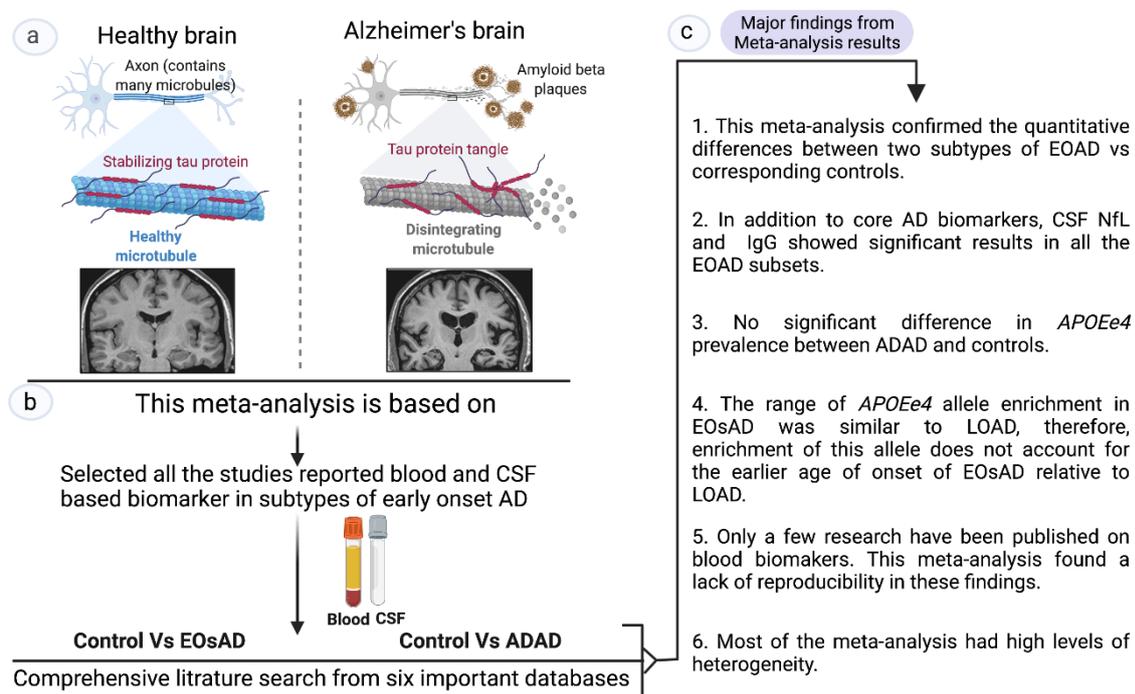


Figure 2.6: A systematic overview of the meta-analysis: (a) The healthy brain has intact neurons and in Alzheimer's brain, tau protein is subjected to various post-translational modifications that reduce the affinity of tau for microtubules and instead, they assemble to form aggregates. Microtubule depolymerization causes loss of axonal integrity, leading to degeneration and eventually cell death, which is anatomically visualized as a shrunken brain in AD cases. (b) This meta-analysis is based on the blood and CSF based biomarkers in subtypes of early onset AD. (c) This section is highlighting the major findings from the meta-analysis and systematic review. This figure was prepared by Ms Kaur using molecular subcomponents provided in Biorender software (Toronto, Canada).

2.4 Discussion

I performed a comprehensive meta-analytic review of all currently available protein/peptide CSF and blood biomarkers of early onset AD subtypes and explored these early onset AD phenotypes in the context of quantified biomarkers. I found that the hallmark proteins ($A\beta$ and tau) showed similar patterns in the CSF, both qualitatively and quantitatively between EOAD and ADAD, which was unanticipated since the expression of pathology, as reflected in classical

biomarkers, might be expected to be less in the EOsAD group, which had no family history of EOAD. Instead, I found that genetic testing ruled out even in the small proportion of APP and PSEN mutations. This outcome suggests that an ADAD-like phenotype, without the accompanying genotype, occurs in EOsAD, a group representing the majority of early-onset cases, but has received surprisingly little attention to date. In addition, other potential emerging markers of neurodegeneration and inflammation, including CSF NfL, IgG and IL-6, were reported in EOsAD. However, the relative paucity of additional biomarkers and low numbers of reported studies was notable, particularly for the relatively accessible serum/plasma sample type.

A β peptide is a hallmark of AD, whose complex lifecycle not only produces a plethora of chain length variants such as A β ₃₈, A β ₄₀, A β ₄₂, A β ₄₃¹⁶ but also multimeric aggregate structures such as oligomers, fibrils and plaque¹⁷. Of these, the 42-amino-acid peptide (A β ₄₂) is the most studied variant due to its enhanced tendency to aggregate⁶⁴. This process, known as the sink model, results in lower levels of soluble A β ₄₂ as plaque/insoluble levels increase¹⁹. Our meta-analysis results, which show a drop in CSF A β ₄₂ in all EOAD variants (Table 2.1), are consistent with this model. These observations are consistent with previous meta-analytic data on LOAD CSF A β ₄₂²¹. Interestingly, the lowest pooled SMD (highest drop in A β ₄₂) was for ADAD symptomatic mutation carriers (sMCs) and the EOsAD subset with no APP, PSEN1 or PSEN2 mutations. The decreased level of CSF A β ₄₂ in ADAD asymptomatic mutation carriers (aMCs) was almost equal to EOsAD and suggested an EOsAD aetiology similar to ADAD. There were few studies of other CSF A β peptides in the context of EOAD. However, a significant reduction in the level of A β ₄₃ in EOAD patients relative to controls has been reported¹⁹².

Tau is one of the most abundant proteins in neurons, with a role in stabilizing the structure of microtubules²⁶. In AD, tau is hyperphosphorylated, dissociated from microtubules and can be detected in CSF at an increased level relative to healthy controls²⁸. Our meta-analyses showed a significant increase in both T-tau and P-tau in all subtypes of EOAD. The pooled SMD of T-tau was significantly higher in ADAD sMCs as compared to EOsAD, likely reflecting a higher degree of neuronal damage in ADAD as compared to the EOsAD variant. The EOsAD subset with confirmed absence of APP, PSEN1/PSEN2 mutations had higher levels of both tau variants than the entire EOsAD group. While the significance of this is not entirely clear, it suggests the possibility that the EOsAD group has mixed aetiologies. Tau can be

phosphorylated on various sites; however, most assays report phosphorylation of thr181 and ser199 specific epitopes because of the higher diagnostic accuracy of these epitopes to differentiate AD from healthy controls²⁹. While our meta-analysis results confirm that these two specific epitopes discriminate between control and AD participants, neither the SMD nor the p-values were higher than the general P-tau measure, suggesting a slight advantage in using these specific phospho-epitopes. In LOAD, several studies have reported other tau epitopes; thr231 was detected before the formation of paired helical filaments²⁹, whereas thr181 and ser199 were considered late events in AD³⁰.

According to some models, neurodegeneration is more closely linked to symptomatology as compared to A β pathology¹⁹³. A handful of other neurodegeneration and inflammation markers, including NfL, albumin, IL-6 and IgG that I identified as suitable for meta-analysis, were reported in only 2 to 4 studies each. CSF NfL was significantly upregulated in EOAD cases compared to healthy controls. NfL is the light chain protein of neurofilament, which makes up neurofilament bundles with counterparts that determine the axonal calibre and conduction velocity²¹. NfL is an integral component of synapses, and loss of synapses may account for the increasing level of NfL in blood and CSF^{194,195}.

Furthermore, recent studies have established a correlation between blood and CSF NfL in late-onset AD¹⁹⁶, making it a potentially helpful biomarker for neuronal apoptosis. Like tau, NfL is associated with the neuronal cytoskeleton, and interestingly, I found a similar increase in the pooled SMD of CSF NfL in EOAD relative to controls (additional data provided in the published article). These findings relating to CSF levels of NfL in EOAD suggest that additional studies in both CSF and serum/plasma are warranted since an additional potential biomarker has been identified, even within this narrow range of replicated markers. In addition, there are several other well studied markers of neurodegeneration, such as sTREM2, VILIP-1, NfH, reported in LOAD^{79,197}, which have not, as yet, been explored in EOAD and ADAD.

A particularly notable feature of our EOAD meta-analysis was the lack of replicated data on blood biomarkers. Blood biomarker studies are more prevalent in LOAD^{21,198}, whereas the EOAD biomarkers were assayed almost exclusively in CSF (Table 2.1). This may reflect clinical practice, whereby CSF may be collected more often in EOAD cases and less frequently in LOAD. Assay of CSF A β and tau for the diagnostic evaluation of dementia has recently been recommended by an international consensus group¹⁹⁹, but the invasive nature of a spinal

tap makes it unlikely that it will be taken up widely²⁰⁰. For plasma A β ₄₂, a comprehensive meta-analysis was performed reporting a significant decrease in plasma A β ₄₂ in clinically diagnosed LOAD cases¹⁹⁸.

In contrast, our meta-analysis of ADAD (all MCs) shows a significant increase in plasma A β ₄₂, likely reflecting the higher peptide level in the plasma due to increased expression of A β ₄₂ or cleavage of APP at this earlier age group, in line with the A β sink model. Previous studies have reported that an increase in plasma level of A β ₄₂ correlates with the risk of AD development. With disease progression, the reduction of the A β ₄₂ level over time has been recorded in plasma, reflecting deposition in plaque^{201,202}. In this younger cohort, plasma clearance may be more efficient. There were insufficient studies on plasma A β ₄₂ in EOAD, so a meta-analysis was impossible. Our meta-analysis on EOAD identified several studies reporting serum/plasma albumin and IgG in which a significant increase of IgG was observed (Table 2.1). These observations reflect findings in previous studies showing increased levels of inflammation mediating proteins and pro-inflammatory IgG, which enhance the inflammatory activity in blood and represent a high-risk factor for AD²⁰³. In ADAD, serum NfL is emerging as a promising biomarker in the literature, but I could not perform a meta-analysis on NfL due to an insufficient number of studies. However, a longitudinal study reported that an elevated level of NfL could differentiate mutation carriers from non-mutation carriers much earlier than absolute NfL levels assayed in cross-sectional studies¹⁸⁹.

Interestingly, discovery-based approaches have the potential to quantify 100s-1000s of proteins. While the majority are unlikely to be significantly changed in AD, the few that do change may provide critical diagnostic tools and insight into disease mechanisms. For example, plasma protein profiling of symptomatic and asymptomatic ADAD patients from the DIAN cohort reported significant differential expression of seven and sixteen proteins in aMC and sMC, respectively. These proteins included complement and inflammation family members, which are known to exacerbate the pathogenic processes of AD²⁰⁴. Promisingly, an essential role for complement proteins in neuronal pruning has recently been identified^{205,206}.

The *APOE ϵ 4* allele is the strongest genetic risk factor for LOAD¹⁵. In LOAD, *APOE ϵ 4* is associated with increased A β levels and risk of A β accumulation at an earlier age, thereby accelerating the clinical onset of AD²⁰⁷. Furthermore, 50% of *APOE ϵ 4* homozygous carriers have been diagnosed AD by 70 years of age¹⁵. Less is known about *APOE ϵ 4* allele carrier

enrichment in EOAD. I identified a significant enhancement in the level of the *APOEε4* allele in EOAD compared to healthy controls: 43% and 30%, respectively. By contrast, the ADAD group had a similar *APOEε4* allele expression level compared to healthy controls, 28% and 29%, respectively, indicating no significant difference (Table 2.1B). The level of reported *APOEε4* allele enrichment in LOAD compared to controls was in the range of 13.7-40%^{13,14}. Our meta-analysis shows EOAD *APOEε4* allele enrichment within the range of that reported for LOAD, and therefore enrichment of this allele does not account for the earlier age of onset of EOAD relative to LOAD. Previous studies indicate a dose-dependent effect of *APOEε4* on CSF Aβ₄₂ level in LOAD²⁰⁸. However, insufficient study numbers were available to explore an *APOEε4* dose effect for EOAD. EOAD cohorts may represent an opportunity to explore novel genetic and lifestyle risk factors responsible for the earlier onset of symptoms in EOAD.

2.5 Limitations

Some of the limitations of this study include the lack of longitudinal data, small sample size cohorts, and fewer studies that include analysis of plasma samples. For ADAD studies, the majority of the meta-analyses were based on two cohorts: DIAN and the Colombian Alzheimer's Prevention Initiative. Considerable heterogeneity was observed across studies, probably due to variation in study characteristics. I managed this in part by using a random-effects model. Only 5 out of our 28 meta-analyses had sufficient study numbers to assess publication bias. The majority of EOAD studies (16 out of 20) did not perform genetic analyses to eliminate APP/PSEN gene mutations and relied purely on the absence of family history. Interestingly, those 4 EOAD studies in which APP/PSEN gene mutations were eliminated had lower levels of Aβ₄₂ and higher levels of all tau variants, suggesting an ADAD-like phenotype without the ADAD genotype, indicating that investigation of additional "non-classical" drivers of EOAD is warranted. Furthermore, study numbers of particular biomarkers (i.e., CSF IL-6, IgG, and albumin) are low, so in some cases, insignificant results could be a reflection of insufficient numbers of published studies, again reinforcing the need for additional biomarker work, particularly for genetically well-characterized EOAD.

2.6 Conclusions

Our meta-analysis of established biomarkers has confirmed the quantitative difference between two subtypes of EOAD vs corresponding control groups. CSF NfL and IgG were statistically significant and emerging as potential markers of neurodegeneration and inflammation to differentiate EOsAD from healthy individuals. EOsAD is a relatively neglected AD subtype of unknown etiology and therefore represents an opportunity to explore as yet unidentified genetic and/or lifestyle factors that contribute to earlier AD onset. To date, the range of biomarkers studied have been heavily focused on CSF, but for establishing promising blood tests, more research is needed before any of these could replace CSF or imaging tests. Generally, EOAD biomarker research is limited to the peptides considered hallmarks of AD pathophysiology, whereas screening a broader spectrum of analytes (e.g., proteomic profiling, lipids, RNA, metabolites) and longitudinal analysis is necessary to track the progress of disease pathology. The studies I included for meta-analysis was based on a dichotomization of age. However, studying age as a continuous variable in cohorts might provide insight into the trajectory of biomarker change with age and stage of the disease. Similarly, a longitudinal study design would offer greater statistical power to identify biomarker changes over time. These are important considerations for the planning and design of new studies. Understanding pathophysiology and identifying additional preclinical biomarkers can be facilitated using discovery-based "omics" approaches.

Chapter 3

Extending the Depth of Human Plasma Proteome Coverage Using Simple Fractionation Techniques

The majority of the content of this chapter has been published:

Gurjeet Kaur, Anne Poljak, Syed Azmal Ali, Ling Zhong, Mark J. Raftery, and Perminder Sachdev. Extending the Depth of Human Plasma Proteome Coverage Using Simple Fractionation Techniques. *Journal of Proteome Research* 2021 20 (2), 1261-1279. DOI: 10.1021/acs.jproteome.0c00670

See Appendix for the complete publication

3.1 Introduction

Human blood, plasma and serum are clinical samples commonly used for diagnostic analyses as they are minimally invasive sample types, already in routine clinical use, and in direct contact with damaged tissues, and therefore frequently reflect downstream effects of disease processes²⁰⁹⁻²¹¹. These sample types contain a diversity of analytes, such as metabolites, lipids and proteins/peptides, of which proteins/peptides are commonly used as biomarkers, since arguably, they represent the full diversity of cellular machinery, are the drivers of all biochemical pathways and are the functional counterpart to the genome. Consequently, they are the most direct and specific cellular and physiological homeostasis or change targets. Immune assays, such as enzyme-linked immunosorbent assay (ELISA), are commonly used for quantitative analysis of individual proteins for disease diagnosis. However, they are inherently targeted approaches helpful in aiding the diagnosis of diseases where the pathophysiology is understood but are of limited use for discovery-based/data-driven investigations, even when used in multiplexed assays²¹². Furthermore, antibody-based approaches frequently lack specificity for proteins and/or isoforms. Consequently, mass spectrometry (MS)-based technology is not only the gold standard for targeted quantification but is the only approach which allows unbiased data-driven investigations. In clinical research, mass spectrometry is, therefore, a method of choice for both absolute and relative quantification of proteomic expression changes, exploration of disease mechanisms and biomarker discovery in plasma²¹³. However, the extended dynamic range of protein abundance in plasma, reported to range between 9-13 orders of magnitude, makes MS-based plasma proteomics extremely challenging^{210,214}.

Over the last three decades, proteomics technologies have evolved and improved so that identification of 100s-1000s of proteins in various tissue samples is now routine. Nonetheless, relatively few biomarkers identified by proteomics techniques have progressed to use in clinical practice²¹⁵⁻²¹⁷. This is partly due to the lack of robust, reproducible, and high-throughput proteomics workflows to determine and validate potential biomarkers in large cohorts.

Some low abundance proteins (LAP) may be transient in the plasma due to variations in the disease state, temporary tissue damage, specific diet and other environmental factors and may

remain undetected. Furthermore, the overwhelming presence of high abundance proteins (HAP), which represent >99% of total plasma protein mass²¹⁸, can obscure lower abundance components from detection by mass spectrometry. Plasma prefractionation methods, therefore, play a crucial role to overcome the complexity of plasma samples, providing the opportunity to identify low abundance proteins²¹⁹. More importantly, tissue-specific proteins, which are more likely to be clinically relevant, are likely to be of medium to low abundance, and therefore are more likely to be identified in fractionated plasma. Various standard plasma prefractionation approaches have been reported in the literature, including affinity enrichment, immunodepletion, and a variety of chromatographic and electrophoretic fractionation techniques. In general, the greater the plasma proteome coverage, the more complex and time consuming the prefractionation workflow²²⁰⁻²²², making most methods impractical for large clinical studies.

The objective of this study was to identify a fractionation strategy which would provide good plasma proteome coverage and identify tissue-specific proteins while being compatible with the higher sample throughput required of many clinical studies. Using a two-step approach, I (1) compared high abundance protein depletion (immunoaffinity columns; Hu6, Hu14) vs low abundance protein enrichment (ProteoMiner combinatorial peptide ligands), followed by (2) further fractionation of the low abundance proteins using chromatographic and electrophoretic approaches (C18 column and 1D SDS PAGE respectively).

3.2 Experimental procedures

To optimize plasma fractionation methodology, whole blood was collected by venepuncture from two healthy volunteers, one male and one female, within the 35-55-year age range. A single pooled sample was used to compare workflows and avoid confounding the method comparisons by biological variation. Blood was collected into EDTA containing tubes, centrifuged (2000g, 20min, 4°C), and the plasma transferred and pooled into a clean 15mL polypropylene tube. To minimize freeze-thaw cycles, plasma aliquots were prepared (250–500 µl) and stored at -80°C till required. The UNSW Human Research Ethics Committee approved a protocol for blood collection. The blood samples used here were collected from volunteers, and both subjects gave written informed consent.

3.2.1 Depletion of high abundant proteins using immunoaffinity-based Columns: Human 6 (HU6) and Human 14 (HU14) columns

I started my experiments with a total volume of 900 µl of plasma which was divided equally into 3 vials (each 300 µl), one for each of the three pre-fractionation techniques (Hu6, Hu14 and ProteoMiner). Depletion of 300 µl of crude plasma using antibody-based immunoaffinity columns HU6 (4.6 x 50 mm, Agilent, CA) and HU14 (4.6 x 100 mm, Agilent CA) was achieved as follows.

3.2.1.1 Plasma preparation for HU6 column:

300 µl plasma was diluted with 1200 µl of Buffer A (1:5 dilution, as recommended by Agilent tech), and each injection was 100 µl on HU6 column (composed of 20 µl plasma+80 µl Buffer A). This step eliminates approximately 85-90% of total plasma protein by binding the six high abundance plasma proteins (albumin, IgG, IgA, transferrin, haptoglobin, and antitrypsin).

3.2.1.2. Plasma preparation for Hu14 column

300 µl plasma was diluted with 900 µl of Buffer A (1:4 dilution, as recommended by Agilent tech), and each injection was 120 µl on HU14 column (composed of 30 µl plasma+90 µl Buffer A). This step eliminates approximately 94% of total plasma protein and contains antibodies to the fourteen high-abundance plasma proteins (albumin, IgG, antitrypsin, IgA, transferrin, haptoglobin, fibrinogen alpha2-macroglobulin, alpha1-acid glycoprotein, IgM, apolipoprotein AI, apolipoprotein AII, complement C3 and transthyretin). Plasma was diluted and filtered using spin filters (Corning Costar Spin-X centrifuge tube filter, 0.45 µm Cellulose Acetate, MERCK, Germany) before chromatography to remove any particulates. Chromatography and fraction collection was performed on an Agilent 1290 UHPLC system (Agilent, Santa Clara, CA). Binding buffer (A) and elution buffer (B) were commercial proprietary products, and the LC methods used were based on the manufacturer's instructions (Agilent, Santa Clara, CA). Highly abundant proteins were retained on the columns while buffer A was in use, and the flow-through proteins represented the low abundance protein fraction. The high abundance proteins were then eluted with buffer B. The low abundance protein fractions were processed for further fractionation and LC-MS/MS analysis.

3.2.1.3 Enrichment of low abundant proteins using bead-based depletion

I sought to compare high abundance protein depletion against a low abundance protein enrichment approach, using beads bound to complex mixtures of combinatorial peptide ligands (ProteoMiner Beads; BIO-RAD, Berkeley, CA). The ProteoMiner bead preparation consisted of 500 μ l bead slurry (20% beads in aqueous ethanol, 20% v/v), resulting in 100 μ l settled bead volume. Plasma (300 μ l) was used for depletion, with the manufacturer manual recommending 200 μ l of plasma sample with a protein concentration of 70 mg/ml. However, our plasma sample had a \sim 50mg/mL concentration, so I adjusted the volume up. The beads and plasma mixture was incubated for 2 h at ambient temperature on a rotation platform. The sample was then centrifuged (1000 \times g, 1 min, ambient temperature), and the flow-through fraction containing the unbound proteins was collected for further analysis. Based on the manufacturer's protocol, I performed three washing steps using a wash buffer, discarding the solution from each wash step. Therefore, the unbound fraction does not include any proteins that may have been present in the wash steps. In addition, this approach reduces the dynamic range of plasma proteins since low and high abundance proteins have a similar chance of being bound to their respective ligands, which are distributed in similar proportions. As a result, excess high abundance proteins are washed away once they have reached their ligand saturation point, whereas low abundance proteins may never saturate their ligand and are enriched within the sample. Both bound and unbound (flow-through) samples were further fractionated (C18 or SDS PAGE) and analysed by LC-MS/MS.

Depleted plasma proteins from the above approaches were concentrated, and the buffer was exchanged to 50 mM ammonium bicarbonate back to the original volume (300 μ l) using Amicon 3kDa concentrators (Millipore, Billerica, MA). Protein concentrations of depleted plasma were determined by use of a nanolitre scale spectrophotometer (DeNovix DS-11+, POCD, Sydney, Australia) used in microvolume mode with 2 μ l of each 3kDa concentrated sample, 280 nm wavelength, and using the extinction coefficient of BSA (E1% (g/100mL) = 6.67). The samples were retrieved from the DeNovix sample surface following absorbance reading to minimize total protein loss. The step was repeated three times, and mean values were used as the final sample concentration. To compare the protein profiles across all methods, 50 μ g of concentrated plasma from all methods were used as the starting point for further fractionation approaches. All subsequent fractionations were each performed in duplicate, providing two technical replicates for each fractionation approach. In this way, the depleted

plasma samples from each of the HU6, HU14, PB and PUB steps (outlined above) was fractionated in duplicate by each of the 1D SDS PAGE and C18 methods (explained below). Therefore, each step generated 24 fractions x2, and all were analysed separately by LCMS/MS using all 8 prefractionation methods, each using an identical pooled aliquot of depleted plasma.

3.2.1.4 Electrophoretic fractionation of low abundance proteins using 1D-SDS PAGE

Equal amounts of protein samples from crude and depleted plasma (50 µg), obtained from low and high abundant fractions of HU6, HU14, ProteoMiner bound (PB) and ProteoMiner unbound (PUB) methods were prepared in duplicate and diluted in 5 µl LDS sample buffer (4x NuPAGE, Invitrogen, USA), 2 µl reducing agent (10x NuPAGE, Invitrogen, USA), and deionized water to make a final volume of 20 µl. Samples were heated for 10 minutes at 70°C, and proteins were separated on NuPAGE 4-12% Bis-Tris midi gels (Invitrogen, USA) using 1X SDS PAGE running buffer and the following electrophoretic parameters; voltage 200V, current 200I for 54 min. (NuPAGE, Invitrogen, USA). Gels were then stained²²³ (colloidal Coomassie G250). After destaining, the separated protein lanes were cut into 24 equal-sized gel bands using a gel cutter tool with a 24-lane blade (Gel Company Inc, CA, USA) and placed into 24 vials for in-gel trypsin digestion.

3.2.1.5 In-gel trypsin digestion

In-gel digestion was performed by adaptation of previously published approaches²²⁴. Prior to performing In-gel digestion, excised gel bands (duplicate gel lanes/depletion method) were destained by incubating in 100mM ammonium bicarbonate/acetonitrile (1:1, 200 µl) overnight at room temp. Proteins were then reduced in an aqueous solution of 100mM ammonium bicarbonate containing 100mM DTT (100 µl) and incubated at 37° C for 1 hour. Proteins were alkylated using 20 µl of 200mM iodoacetamide (IAA) and incubated at 37° C for 30 min. The reduction/alkylation solution was decanted, and the gel bits were further washed using 100% acetonitrile to remove DTT and IAA reagents. Trypsin digestion (Promega, sequence grade modified trypsin, USA) was performed (10 ng/2 µl) in 100 µl 10mM ammonium bicarbonate at 37° C for ~18hrs. Formic acid (0.1%, 20 µl) was added and incubated at ambient temperature (10 min.), followed by the addition of 100% acetonitrile (50 µl). The extracted peptide solution was transferred to LC vials (Agilent Technologies, USA) and dried using a centrifugal evaporator (SpeedVac, Thermo Fisher Scientific, USA). Peptides were redissolved in 5 µl of

0.1% heptafluorobutyric acid containing 1% formic acid and were then ready for LC-MS/MS analysis.

3.2.1.6 In-solution protein digestion

For chromatographic peptide fractionation, proteins were digested *in situ* using 50 µg of initial protein sample from all depletion methods in duplicate. Proteins were reduced by adding 2 µl of tris-(2-carboxyethyl) phosphine (TCEP) and incubated for 60 min at 60° C. Proteins were further alkylated using 1 µl of 200mM iodoacetamide (37mg IAA/mL), vortexed, spun briefly and incubated at ambient temperature for 10 min. The sample pH was checked and adjusted to pH 7-9 if necessary, then trypsin was added (4.4 µg/10µL) (Promega, sequence grade modified trypsin, USA) and incubated at 37° C for ~16hrs. The extracted peptide solution was transferred to LC vials (Agilent Technologies, USA) and dried using a centrifugal evaporator (SpeedVac, Thermo Fisher Scientific, USA). Peptides were redissolved in 5 µl of 0.1% heptafluorobutyric acid containing 1% formic acid and were then ready for LC-MS/MS analysis²²⁵.

3.2.1.7 Chromatographic fractionation of low abundance proteins using a C18 column (ZORABX extended-C18)

A binary gradient achieved the chromatographic separation of tryptic peptides; Buffer A consisted of 20mM ammonium formate containing 2% acetonitrile (pH 10), and buffer B consisted of 20mM ammonium formate containing 90% acetonitrile (pH 10). The 64.5 min gradient was started with a 5 min isocratic step of 100% buffer A, followed by several gradient steps; 0-15% buffer B in 8 min; 15-28.5% buffer B in 33 min; 28.5-34% buffer B in 5.5 min; 34-60% buffer B in 13 min. Peptide fractions were collected at 0.6 min time intervals, resulting in a total of 84 fractions. These were then concatenated by pooling equally spaced early, mid and late fractions, and the pattern repeated to a final number of 24 peptide fractions. The concatenation resulted in fractions consisting of peptides across the entire chromatographic timespan. The fractions were dried by vacuum centrifugation (SpeedVac, Thermo Fisher Scientific, USA), re-suspended in 1% formic acid containing 0.1% HFBA, and were ready for LC-MS/MS.

3.2.2 Mass Spectrometry

Technical duplicates of each sample were analyzed using a Q-Exactive Plus mass spectrometer (Thermo Electron, Bremen, Germany), using an adaptation of previously published work²²⁶. Peptide separation was carried out using a nano-LC, Dionex UltiMate 3000 HPLC system (Thermo Scientific, Waltham, USA), equipped with an autosampler (Dionex, Amsterdam, Netherlands). Peptides (~2µg on-column) were initially captured onto a C18 cartridge (Acclaim PepMap 100, 5µm 100 Å, Thermo Scientific Dionex, Waltham, USA), switching to a capillary column (25 cm length, 350 µm o.d., 75 µm i.d.) containing reverse phase packing (C18, Reprosil-Pur, 1.9 µm, 200 Å, Dr Maisch GmbH, Ammerbuch-Entringen, Germany), placed within a column heater (45°C, Sonation GmbH, Germany). Peptide elution involved a 60 min run time and a binary gradient of 0 - 45% buffer B, at 200 nL/min. The binary buffers consisted of buffer A (H₂O: CH₃CN of 98:2 containing 0.1% formic acid) and buffer B (H₂O: CH₃CN of 20:80 containing 0.1% formic acid). Mass spectrometer settings were: ion spray voltage 2000V, capillary temperature 275–300°C, positive ion mode. The method is a shotgun sequencing approach, using data-dependent acquisition (DDA), with a survey scan acquired (m/z 375-1750) and up to ten multiply charged ions (charge state $\geq 2^+$) isolated for MSMS fragmentation (intensity threshold of 8.0×10^4 , with nitrogen as HCD gas). The auto MS/MS acquisition consisted of 10 scans per duty cycle, MS resolution of 30,000 and MSMS resolution of 70,000.

3.2.3 Computational Analysis

Initially, the data processing of the raw files was performed on MaxQuant (version 1.6.10.43)²²⁷ using the Andromeda search engine²²⁸. I used the filter of ≥ 2 unique peptides per protein for profiling and quantitation. The UniProt *Homo sapiens* database (Proteome ID UP000005640, Swiss-Prot and TrEMBL) was used with reverse decoy protein sequences (used for protein identification false discovery rate estimation) frequent contaminant sequences. The default value of precursor mass tolerance (± 0.07 Da) for the first search and (± 0.006 Da) for the main search was applied. Mass tolerance for matching peaks to theoretical MSMS ion series was five ppm. Protein identification false discovery rate (FDR) was set to <1% to ensure only high-confidence identifications for the peptide to spectra match, protein, and site decoy fractions. Enzyme specificity was set to trypsin, with a maximum of two missed cleavages. Searches included variable modifications of protein N-terminal acetylation, methionine

oxidation, pyroGlu and pyroGln and fixed modification of carbamidomethylation of cysteines. The search tolerance used for both peptide precursor match and fragment ions was 20 ppm, and only those peptides of more than 6 amino acid residues in length. Peptide spectral matches (PSM), peptides and protein were selected based on a false discovery rate (FDR) of less than 0.01 based on a reversed sequence decoy database. Furthermore, only peptides with ≥ 2 PSM were selected following filtering with these criteria. To validate and transfer identifications across adjacent fractions and technical replicates, the 'match between runs' (MBR) option in MaxQuant was enabled, with a retention time window of 0.7 min and an alignment time window of 20 min.

All the commonly identified proteins in all methods were quantified for peak area abundance within the MaxQuant environment using the MaxLFQ (label-free quantification) algorithm²²⁹. The MaxLFQ algorithm combines the measured intensities from all matched peptides combined into proteins for final abundance data²³⁰. Finally, the MaxLFQ values obtained from MaxQuant for each protein were statically analyzed in the Perseus environment. This workflow of Andromeda \rightarrow MBR \rightarrow MaxLFQ \rightarrow Perseus is frequently used in studies using data-dependent quantitative proteomics²³¹.

The raw files were also processed through the Trans-Proteomic Pipeline (TPP) version 5.1.0, released on 2017-11-03. This software considers additional factors, including the number of sibling experiments identifying the same peptide ions, the number of replicate ion identifications, sibling ions, and sibling modification states. TPP uses the iProphet algorithm to determine protein identification's false discovery rate (FDR). To perform the TPP analysis, the LC-MS/MS raw files were first converted to open format mzML files using the MSconvertGUI with the default parameters. The Uniprot database and sequence input were similar to the MaxQuant search. The 384 LC-MS/MS raw files were processed using three search engines; X! Tandem (with the *k*-score plug-in)²³², Mascot and Comet, with all parameters kept similar. The PeptideProphet and ProteinProphet algorithms were employed to compute the probability scores for peptide and protein identification. The accurate mass model in PeptideProphet was used for high confidence peptide identifications²³³ and a requirement that each protein is identified with a minimum of two peptides, each with a peptide probability score $>99\%$. All three search engine results were merged and validated using iProphet²³⁴. An iProphet probability of >0.99 was used as the cutoff for final protein identification²³⁵. The total sum mapping identified 10 million spectra, but I used only the common peptides representing

only the highest confidence data and without a search engine bias. Prior to further data processing of our confidently identified protein lists, contaminant proteins and a few remnant isoforms (mainly resulting from unreviewed UniProt TrEMBL data) were removed manually.

3.2.4 Density and Multi scatter plots

All analyses were performed using a combination of MS Excel (<https://www.microsoft.com/en-us/microsoft-365/excel>), R-language (<https://www.r-project.org/about.html>) and Perseus (<https://maxquant.net/perseus/>) platforms. For density scatter plots, and multi scatter plots using `gplots` and `ggplot2`, packages in R were employed. The correlation was assessed for replicating LFQ intensities using the Pearson correlation coefficient. Multi scatter plots are prepared by combining all of the experiments and calculating the Pearson correlation of the expression data across methods.

3.2.5 Hierarchical Clustering (HCA)

Pairwise comparison of sample distance and similarity analysis was performed using Hierarchical Clustering (HCA). The KMeans clustering method was used for all protein LFQ intensities and was performed using the Heatmap package in R. The “one minus Pearson correlation coefficient” with the average linkage method was employed. Boxplots were prepared using the `ggplot2` package in R.

3.2.6 Volcano plot and Principal Component Analysis (PCA)

To compare depletion and fractionation approaches with the aim of identifying any significant protein fold change across methods, I used Volcano plots and principal component analysis (PCA). A two-sided student's T-test for p-value calculation was used for the volcano plot. The test parameter includes the minimum number of 9 valid values in full mode (all groups per row). Correction for multiple testing was performed using the Benjamini-Hochberg correction, with an FDR of less than 0.05 accepted as significant. I used the KEGG database for pathway annotation of the whole plasma proteome, using all 28 pairwise method comparisons, encompassing all 3,548 proteins which overlapped across methods. Significant differences in protein recovery/identification across methods were accepted with a $-\text{Log } P$ value >2.5 , and these were represented on volcano plots by differential colouring (yellow dots). The PCA analysis (plotted in the Perseus environment) was used to determine whether the separation methods and technical replicates would be differentiated globally.

3.2.7 Human Protein Atlas Mapping

I used data in the Human Protein Atlas database (<https://www.proteinatlas.org/>) and the 3,597 confidently identified proteins in our plasma proteome, which also overlapped across methods to determine the tissue specificity of each protein. First, confidently identified proteins were manually curated against the Protein Atlas lists, identifying tissue expression levels across the 51 tissues and/or tissue fractions reported in The Human Protein Atlas. This was done to establish the level of tissue specificity of each of the plasma proteins identified. Then, the data was collated to represent proteins specific to the 11 human organ systems (digestive, circulatory, integumentary, nervous, urinary, skeletal, lymphatic, respiratory, muscular, endocrine, and reproductive).

3.3 Results

3.3.1 Comparative proteomic data provided by plasma depletion and separation workflow

The 2-step plasma fractionation approaches that I implemented in this study (Figure 3.1A) resulted in 8 sample types (Figure 3.1A), each with 24 fractions and 2 technical replicates, which were analysed by LC-MS/MS and resulted in 384 raw files, and a total of 2,890,105 spectra (Table 3.1), of which approximately 25% were assigned to peptides following processing with search engines. All raw data files have been uploaded to ProteomeXchange (<http://www.proteomexchange.org/>) with identifier PXD02246. Comparison of the high abundance protein fraction and the depleted plasma SDS PAGE profiles showed that substantial amounts of the higher abundance proteins were extracted into the Hu6, Hu14 and PUB fractions, and the low abundant protein fractions showed a much more complex banding pattern, in comparison to undepleted plasma (Figure 3.2). Using computational bioinformatics analysis in MaxQuant: search engine, Andromeda, and Trans Proteomics Pipeline (TPP) search engine, Mascot, Comet, XTandem, (supplementary data presented in the full paper in the appendix). I found that the commonly mapped transitions from all four search engines (Andromeda, Mascot, Comet, XTandem) resulted in the identification of 90,237 peptides and ~4,800 high confidence proteins (Table 3.1). In addition, an average of 8,845 peptides per fraction and 22,560 peptides per depletion strategy were identified. A schematic representation of the workflows used for methods development is shown in Figure 3.1, and the results are summarized in Table 3.1.

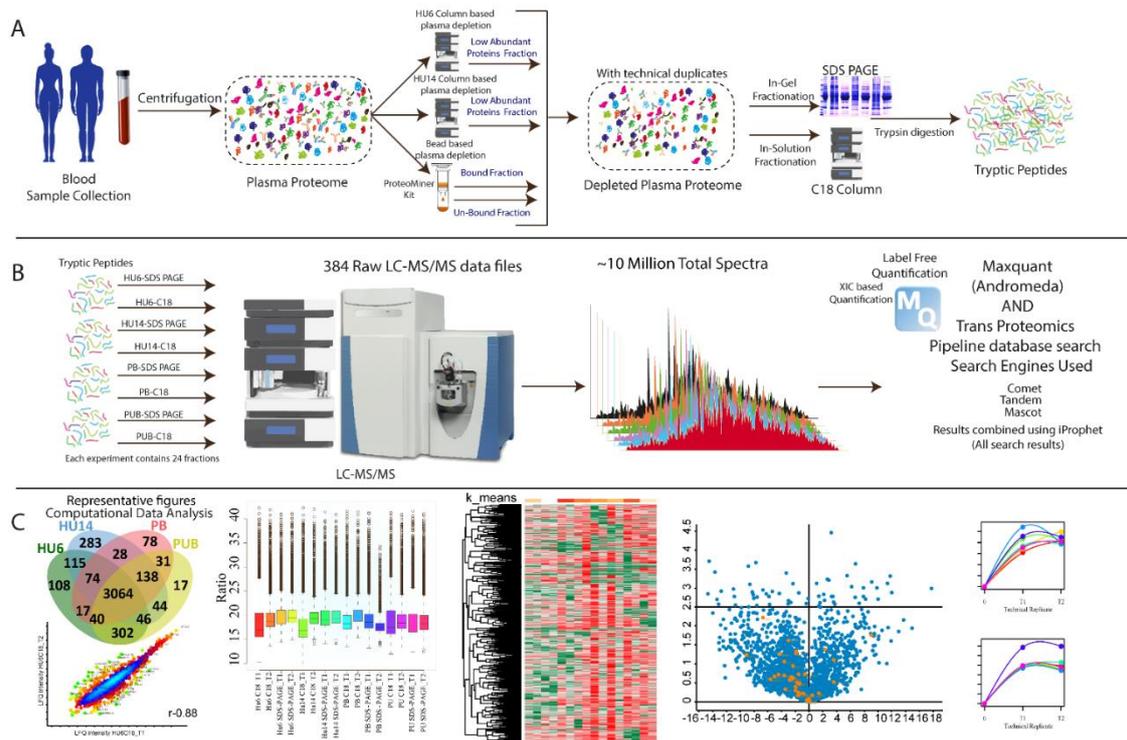


Figure 3.1. Schematic representation: workflow used to optimise plasma fractionation and proteomics analysis. (A) Fractionation approaches, (B) LC-MS/MS analysis and data processing. (C) Computational analysis.

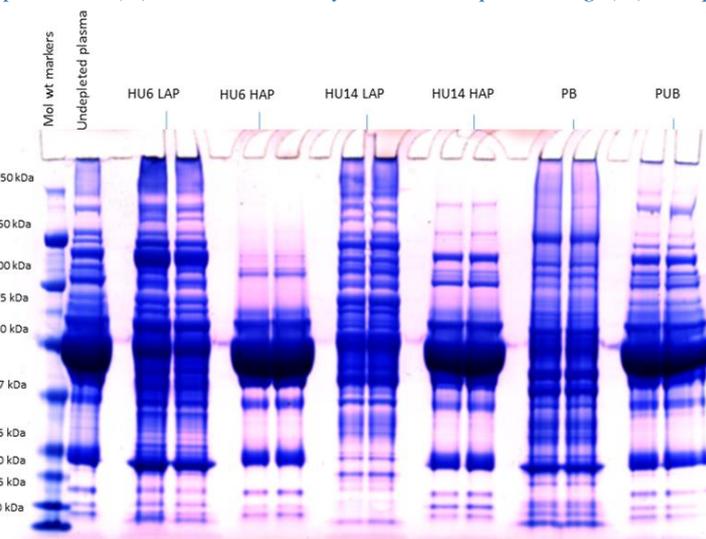


Figure 3.2. Representative NuPAGE LDS gel profile of undepleted plasma and the depleted plasma contains low abundant proteins (LAP) and high abundant proteins (HAP) plasma proteome using HU6, HU14 and ProteoMiner depletion methods (50 µg total depleted proteins were loaded per gel lane).

Table 3.1: Equivalent amounts of total plasma low-abundance proteins (50 µg) were prepared by three different methods of eliminating high-abundance protein components (Hu6, Hu14, and ProteoMiner) and two fractionation approaches (SDS and C18). Two technical replicates were acquired per sample (T1 and T2, with T1T2 representing a database search with combined technical replicates). Summarized data on numbers of spectra, peptides, and proteins acquired and identified are shown, including the total number of query spectra (all LC-MS/MS spectra acquired, including background and peptide signals), assigned spectra (these are the spectra for which a protein database match was identified), and identified peptides (these are the unique plus razor peptides identified). They are fewer than the assigned spectra because frequently, multiple spectra for the same peptide are acquired due to the chromatographic peak width often exceeding the delay time set before the same mass precursor is again subject to LC-MS/MS, total proteins (the proteins identified for each fractionation approach and each workflow), and the common proteins (proteins that overlap across all workflows).

Work Flow	Peptide fractionation approach	Number of Fractions	Technical Replicate	Total Number of queries Spectra	Assigned Spectra	Total peptides	Total Proteins per Fractionation Approach	Total Proteins	Common proteins
HU6 (high abundance protein removal)	SDS	24 Fractions	T1	279069	90752	13132	4209	4209	3597
			T2	282075	82494	12355	4182		
			T1T2	561144	173766	15080	4219		
	C18	24 Fractions	T1	111445	27656	8238	3696		
			T2	148551	31216	8099	3652		
			T1T2	259996	59226	10950	3680		
HU14 (high abundance protein removal)	SDS	24 Fractions	T1	217542	54719	11536	4186	4252	
			T2	184648	57803	12139	4125		
			T1T2	391084	118476	14971	4190		
	C18	24 Fractions	T1	103220	26985	8068	3819		
			T2	107281	22359	6241	3916		
			T1T2	210501	49816	9944	4010		
Proteominer Bound	SDS	24 Fractions	T1	233641	73377	9949	2813	3886	
			T2	217027	61341	9164	2974		

Extending Plasma Proteome Coverage

Chapter 3

(low abundance protein enrichment)			T1T2	450668	134718	11239	3254	4285
			T1	94606	19451	5567	3406	
			T2	90944	23686	7064	3812	
	C18	24 Fractions	T1T2	185550	43645	8663	3825	
Proteominer Unbound (low abundance protein enrichment)			T1	177590	45221	7774	4191	
			T2	204826	49603	7858	3840	
	SDS	24 Fractions	T1T2	382416	95521	9867	4198	
			T1	217538	47400	6507	4009	
			T2	220102	51204	7820	4191	
	C18	24 Fractions	T1T2	437640	99296	9523	4211	

The highest number of total proteins (4,219) were identified with HU6 depletion SDS PAGE fractionation, and a similar number of proteins (4,190) were identified with the HU14 depletion SDS PAGE fractionation, and a similar number of proteins (4,190) were identified with the HU14 depletion SDS PAGE fractionation approach. The lowest number of proteins (3,297) were identified in the PB low abundance proteins enrichment SDS PAGE fractionation method (Table 3.1). Of the ~4385 total proteins confidently identified, the majority are common to all methods, i.e., 3064, representing ~70% overlap across methods (Figure 3.3A), while the approximately 3000 were common in both SDS PAGE and C18 fractionation techniques (Figure 3.3B and C). The detailed information for all the identified LFQ intensities of 3064 proteins were extracted and used for downstream analysis (supplementary data presented in the full paper in the appendix).

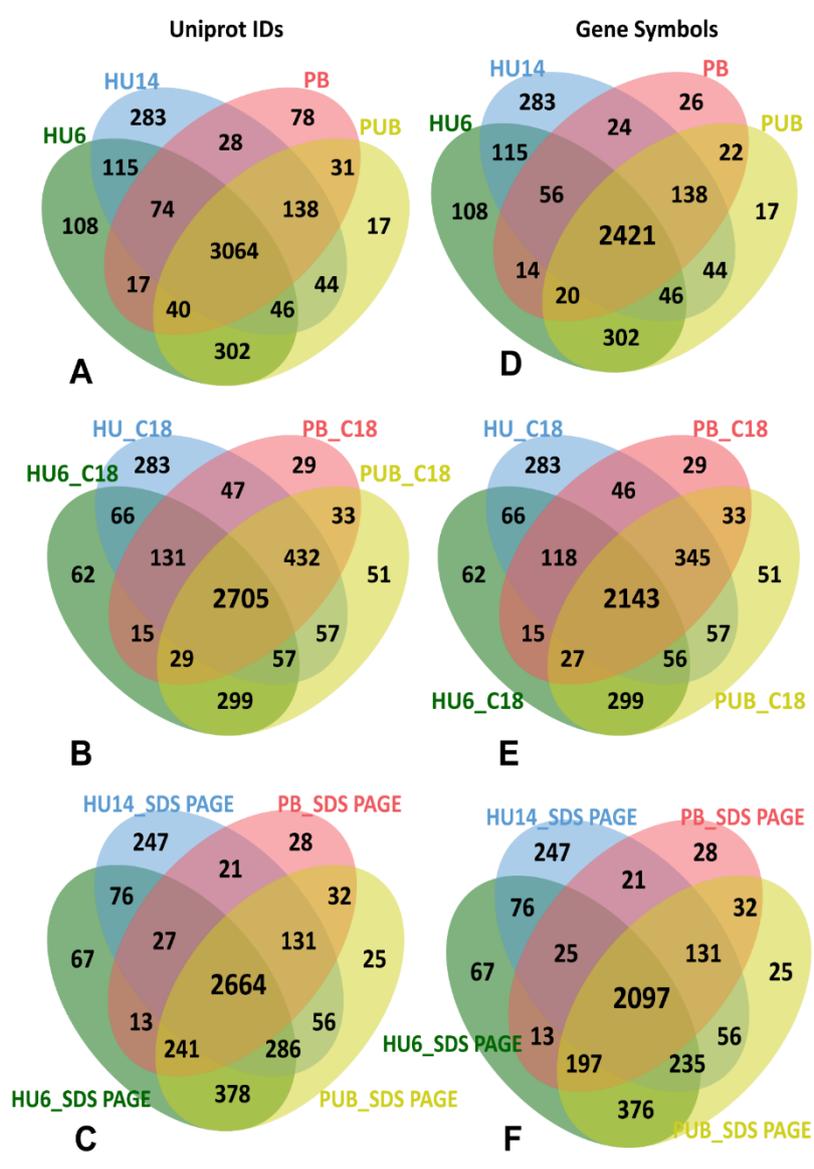


Figure 3.3. (A) Venn diagram shows the protein overlap across all depletion methods (HU6, HU14, PB, and PUB) for all proteins identified with unique Uniprot accession numbers. (B) Method comparison of proteins with unique Uniprot accession numbers using the C18 fractionation approach. (C) Method comparison of proteins with unique Uniprot accession numbers using the SDS PAGE fractionation approach. (D) Venn diagram showing protein overlap across all depletion methods together (HU6, HU14, PB, and PUB) for proteins with both unique Uniprot accession numbers and gene symbols. (E) Comparison of proteins identified with unique Uniprot accession numbers and gene symbols using the C18 fractionation approach. (F) Comparison of proteins identified with unique Uniprot accession numbers and gene symbols using SDS PAGE fractionation.

3.3.2 Technical performance across workflows

Technical replicates were highly correlated across all eight methods, with an average Pearson correlation coefficient of $r = 0.90$, and density scatter plots of technical replicates are shown in Figure 3.4. While the within method technical replicate correlations are high, the cross-method correlations are lower. A detailed description is provided in the multi-scatter correlation plot (Figure 3.5).

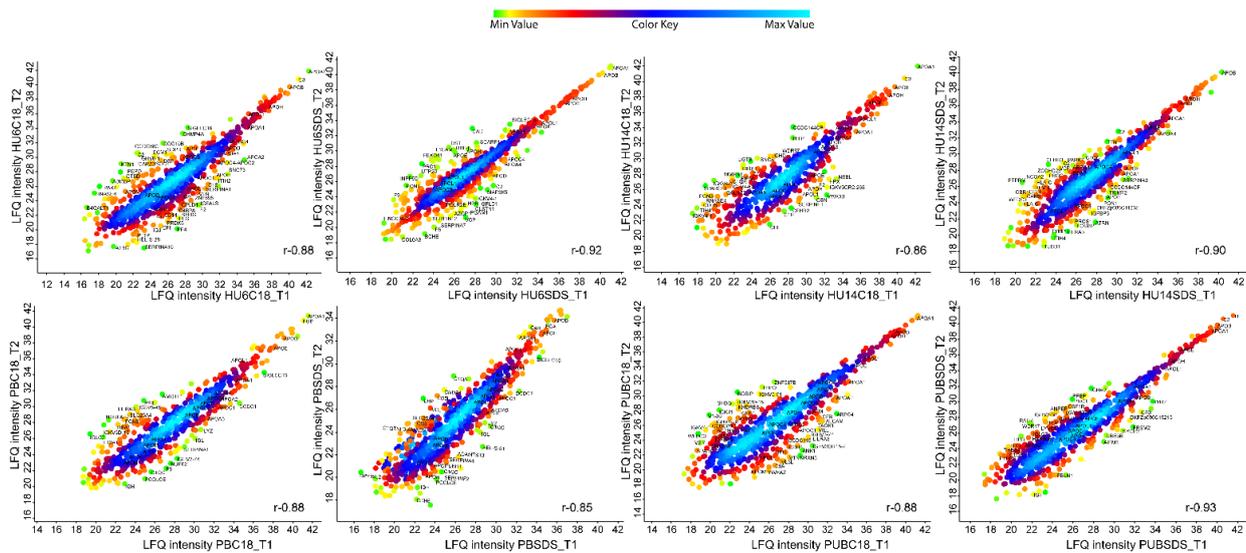


Figure 3.4. Density scatter plots of technical replicates for each of the eight methods, with a density color scale showing areas of higher and lower data density (blue and green respectively) Pearson correlation are indicated as r values on each plot.

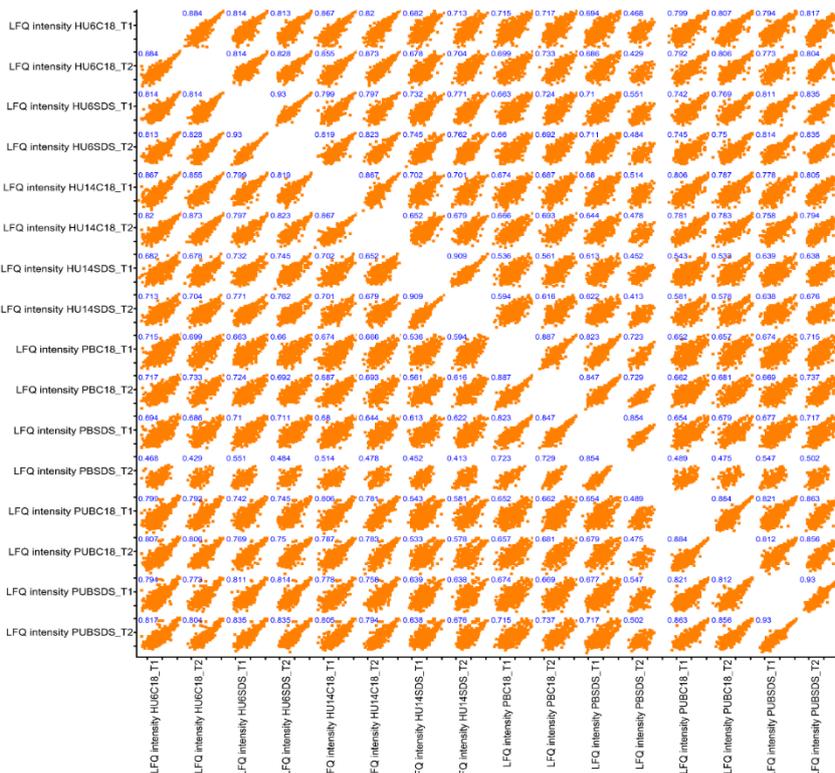


Figure 3.5. Multiscatter plot depicting the cross-correlation of all the methods among each other.

Box plots showing the distribution and variations of the LFQ intensities in all 16 analyses (8 methods, each with 2 technical replicates Figure 3.6A. The line within the box denotes the median value, the upper and lower ranges of the box indicate the 5 and 95 percentiles of the intensities. The dots above the box plots are outliers). Hierarchical clustering analysis (HCA) of the 16 analyses resulted in the formation of the 8 clusters, with associated technical replicates grouped most closely, as shown in Figure 3.6B. Dendrogram based hierarchical relationship (Figure 3.6B) and similarity matrix analyses (Figure 3.6C) showed the close association of protein expression data between the following method pairs: HU6 SDS PAGE and HU14 SDS PAGE, HU6 SDS PAGE and PB SDS PAGE, HU6 SDS PAGE and PUB SDS PAGE, HU14 SDS PAGE and PB SDS PAGE, PB SDS PAGE and PB C18.

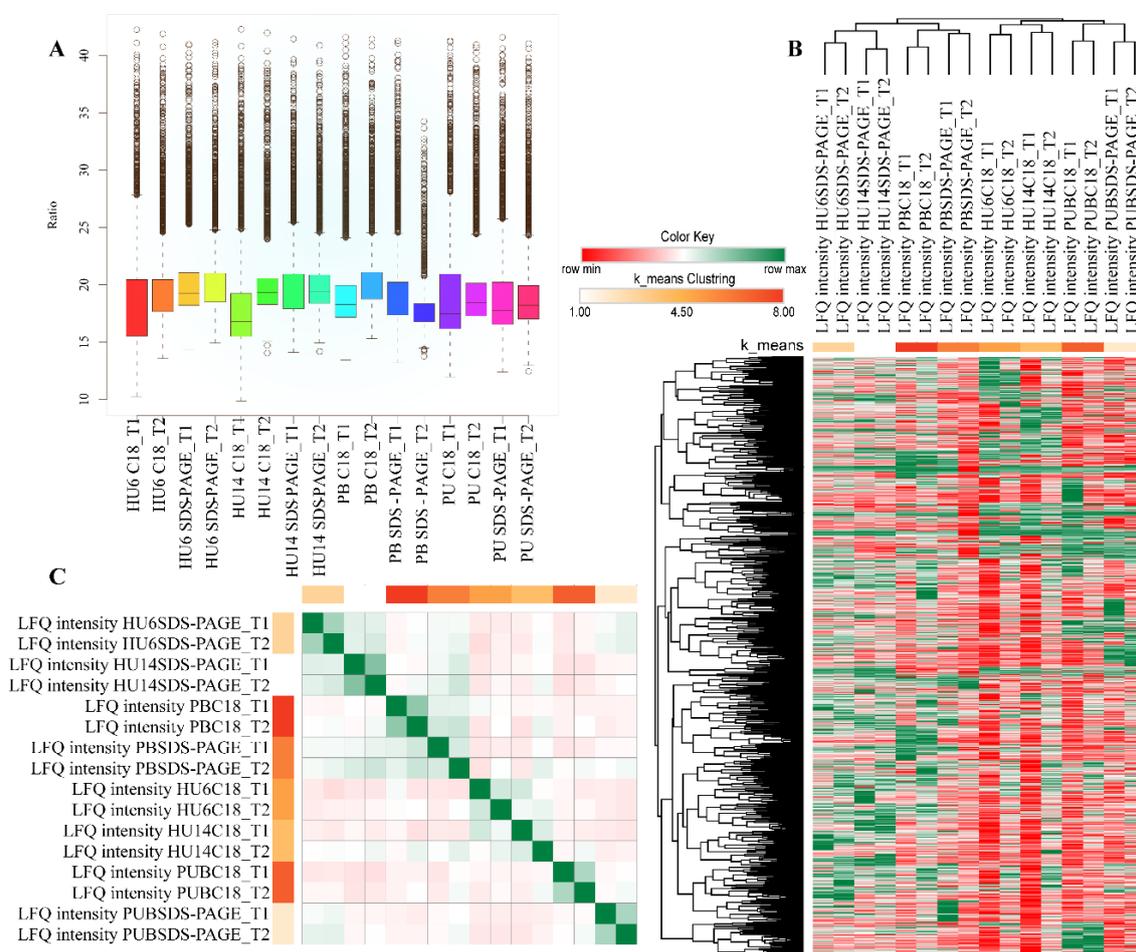


Figure 3.6. A. Box-dot plots of LFQ intensity of all 16 experimental runs. The line within the box denotes the median value, the upper and lower ranges of the box indicate the 5 and 95 percentiles of the intensities. The dots above the box plots are outliers; B. Hierarchical Cluster analysis and heat map for the 3597 proteins common to all methods. The k means algorithm was used to identify protein clusters. C. The similarity matrix and heat map were constructed using the Pearson correlation values of the 16 methods, clustered based on the k means algorithm.

3.3.3 Comparison of different pre-fractionation and separation techniques

I then compared the LFQ intensities merging the technical replicates to compare the C18 and SDS PAGE separation techniques (Figure 3.7). In each case, the relative peak intensities (LFQ intensity) profile is shifted to the right for the SDS PAGE relative to the C18 fractionation methods, indicating greater peak intensity/abundance in the SDS PAGE methods. This observation is also recapitulated in Table 3.1, which shows higher numbers of spectra, peptides and proteins for most SDS PAGE methods compared with C18 methods. Almost equal LFQ intensities were observed in the PB and PUB fractions, indicating that a substantial number of proteins were present in both fractions. The density overlay also shows similar signal intensities of the proteome, indicating that the number and intensity of proteins substantially overlap in the PB and PUB fractions.

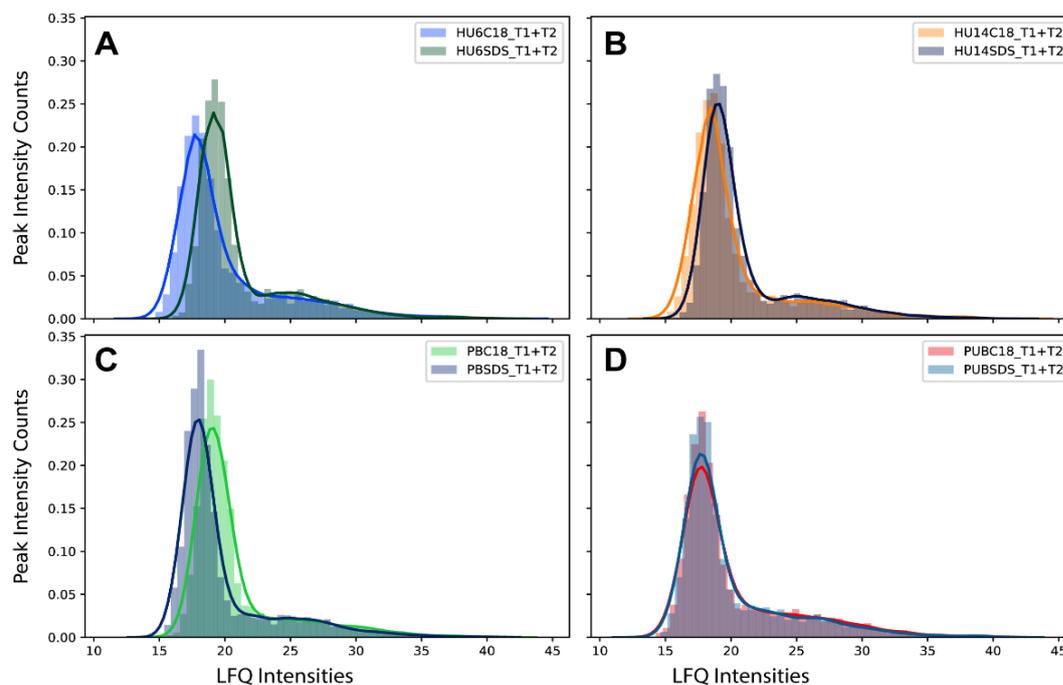


Figure 3.7. Juxtaposed LFQ intensities for pair-wise comparison of C18 vs SDS PAGE fractionation using the following initial depletion approaches; **A.** HU6 **B.** HU14 **C.** ProteoMiner Bound (PB) **D.** ProteoMiner Unbound (PUB).

3.3.4 Principal Component Analysis (PCA)

PCA was performed to explore the degree of differentiation or similarity between technical replicates and across methods. The first and second principal components explained 47.8% and 12.9% of data variability, respectively. The PCA results show that the technical replicates

cluster together as might be expected, and all methods are discriminated from one another (Figure 3.8A). However, closer positioning of specific methods on the PCA plot is observed, particularly in the principal component 1 (PC1) dimension, for the following method pairs; HU6 SDS PAGE and HU14 SDS PAGE, HU6 C18 and HU14 C18, and PB C18 and PUB C18 (Figure 3.8A). A sigmoidal-shaped plot was obtained when I ranked proteins based on the PC1 score (Figure 3.8B), with a long flat section extending across most proteins (*ca* protein rank 200 – 3500). This indicates that variability across methods was constant for the majority of proteins. In contrast, a few proteins had extremely low variability across methods (curve tail facing upwards on the far right), and a few proteins had higher variability across methods (curve tail facing downwards on the far left). Members of two large families of proteins, apolipoproteins and the serpin proteases, are marked on the PC1 plot (Figure 3.8B), and I note that the level of PC1 variability across methods seems to be a function of protein concentration; the higher concentration plasma proteins (APOA2, APOC4-APOC2, APOC1, SERPINA1, SERPINF2, SERPINA6) are detected at the higher rank (right side of the curve), and lower concentration plasma proteins (APOH, APOB, APOA1, SERPINB1, SERPINE1, SERPINA10) appear in the lower rank (left side of the curve) Figure 3.8B. I also ranked proteins for each depletion method based on their LFQ intensities together and separately (Figure 3.10) and annotated the plots with 9 proteins whose protein concentration in plasma is known. This concentration range spans from low pg/mL to low mg/mL, representing ≥ 9 orders of magnitude of concentration. Further, I compared the top 100 dynamic range of proteins in all depletion methods using Venn diagrams to find the number of overlapping proteins in C18 and SDS PAGE, and I found $\geq 70\%$ protein overlap except HU14, which showed 57% (Figure 3.9).

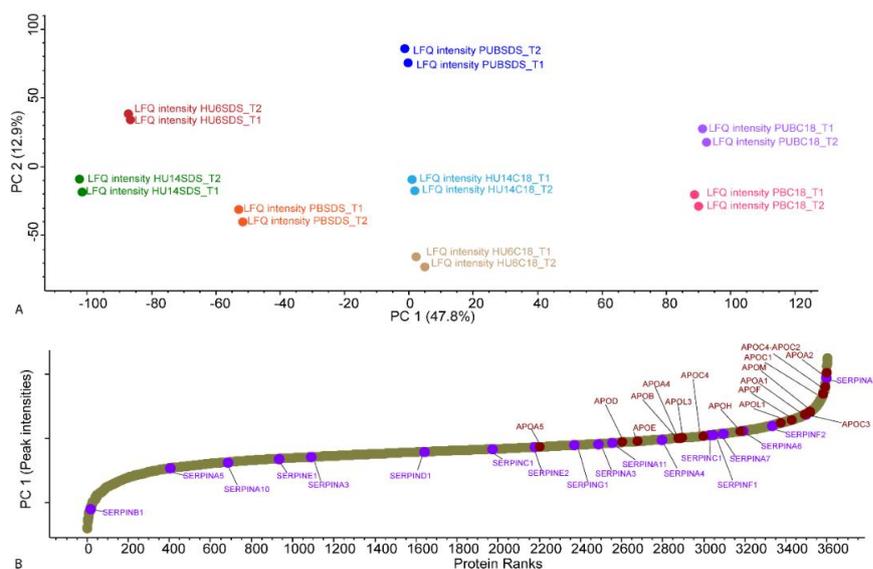


Figure 3.8. Principal Component Analysis: A. The 2D principal component analysis uses the first two principal components B. 1D classification of all the proteins identified in protein rank order based on the first principal component. The long flat section in the sigmoidal shaped plot indicates that the variability across methods was constant for the majority of proteins. The level of PC1 variability across

methods seems to be a function of protein concentration, with higher concentration plasma proteins of apolipoproteins and the serpin proteases family (APOA2, APOC4-APOC2, APOC1, SERPINA1, SERPINF2, SERPINA6) detected at the higher rank (right side of the curve) and lower concentration plasma proteins (APOH, APOB, APOA1, SERPINB1, SERPINE1, SERPINA10) appearing in the lower rank (left side of the curve) marked APOE in red and SERPIN in purple colour.

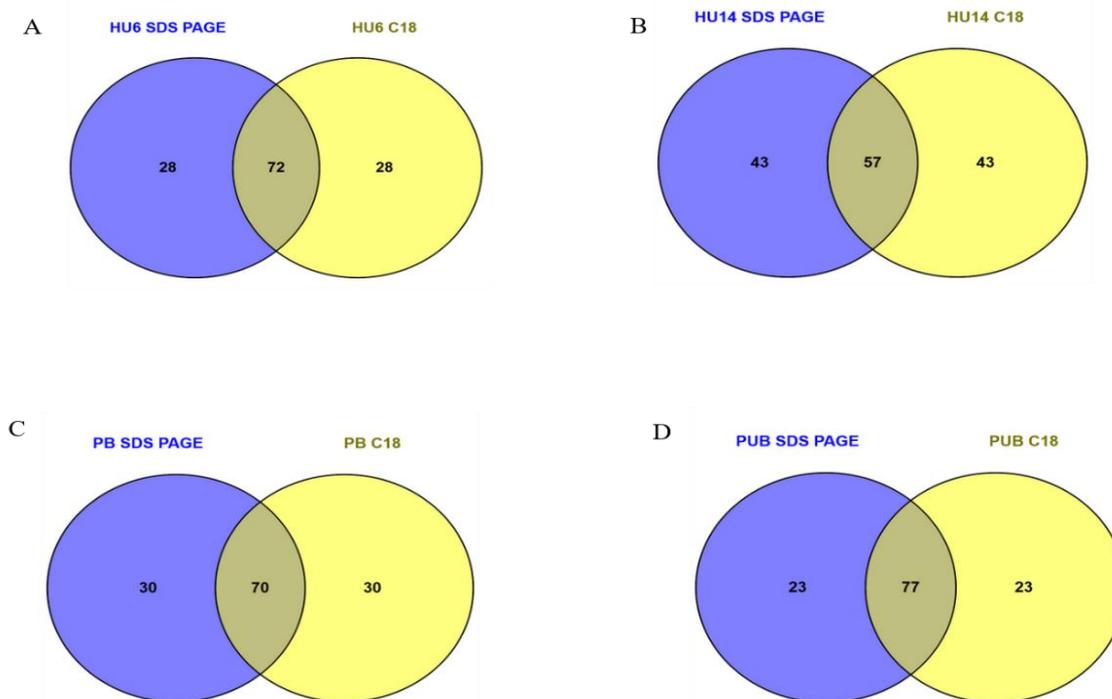


Figure 3.9. Venn diagram showing overlap of top 100 dynamic range of protein between depletion methods A. HU6 C18 and HU6 SDS PAGE, B. HU14 C18 and HU14 SDS PAGE, C. PB C18 and PB SDS PAGE D. PUB C18 and PUB SDS PAGE.

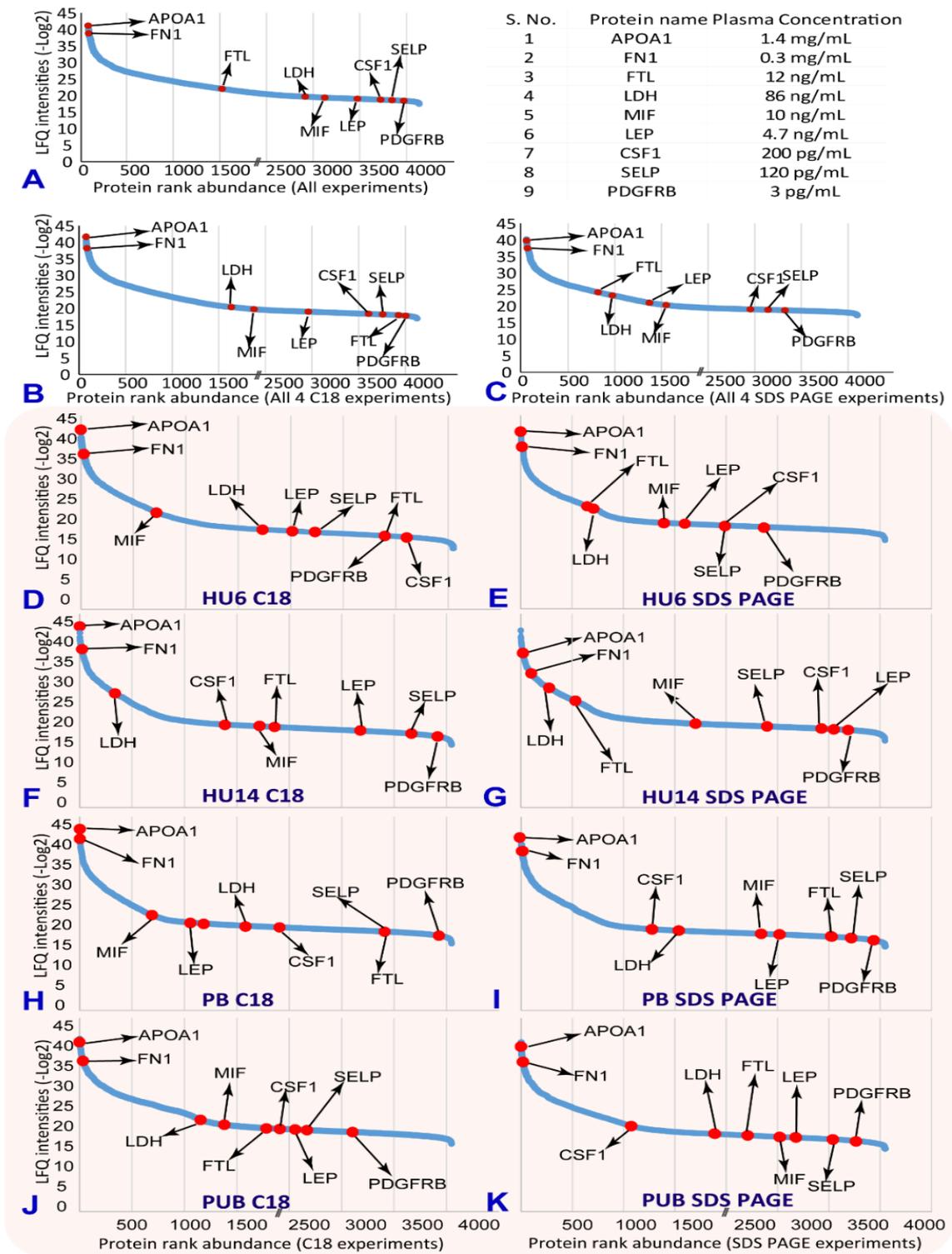


Figure 3.10. Protein rank abundance plot based on LFQ intensities. **A.** All experiments together **B.** the four SDS PAGE methods. **C.** the four C18 fractionation methods, **D.** HU6 C18, **E.** HU6 SDS PAGE, **F.** HU14 C18, **G.** HU14 SDS PAGE, **H.** PB C18, **I.** PB SDS PAGE, **J.** PUB C18, **K.** PUB SDS PAGE, are presented in plots separately. The plot's x-axis describes the rank of the proteins, while the y-axis denotes Log2 transformed LFQ intensities. Nine proteins whose protein concentration is identified in the literature are marked on the plots, and their concentrations are reported in the bottom-right table.

3.3.5 Volcano plot analysis

To evaluate possible systematic bias across methods tested, I performed a volcano plot analysis for all combinations of methods, resulting in a total of 28 plots (Figure 3.11). I compared all the methods to identify the variations in the statistically significant proteins using common proteins across methods. Using the 3,548 proteins common across methods (Figure 3.3), the UniProt-based KEGG pathways identified a total of 39 proteins that were significantly differentially abundant across methods (Figure 3.11), representing ~1% of the full plasma proteome identified here. This slight overall difference across methods is not too surprising, given that an identical sample was analysed across workflows, and indicates that significant quantitative differences attributable to the different sample workup methods are minimal. Furthermore, this number can be taken as a quasi-indicator of protein quantification false discovery rate, and at ~1%, it is encouragingly low. One of these proteins is complement C3, a protein depleted using the Hu14 column, and it is of interest that this is picked up as one of the 39 differentially expressed proteins since the difference, in this case, is likely real and attributable to the method used.

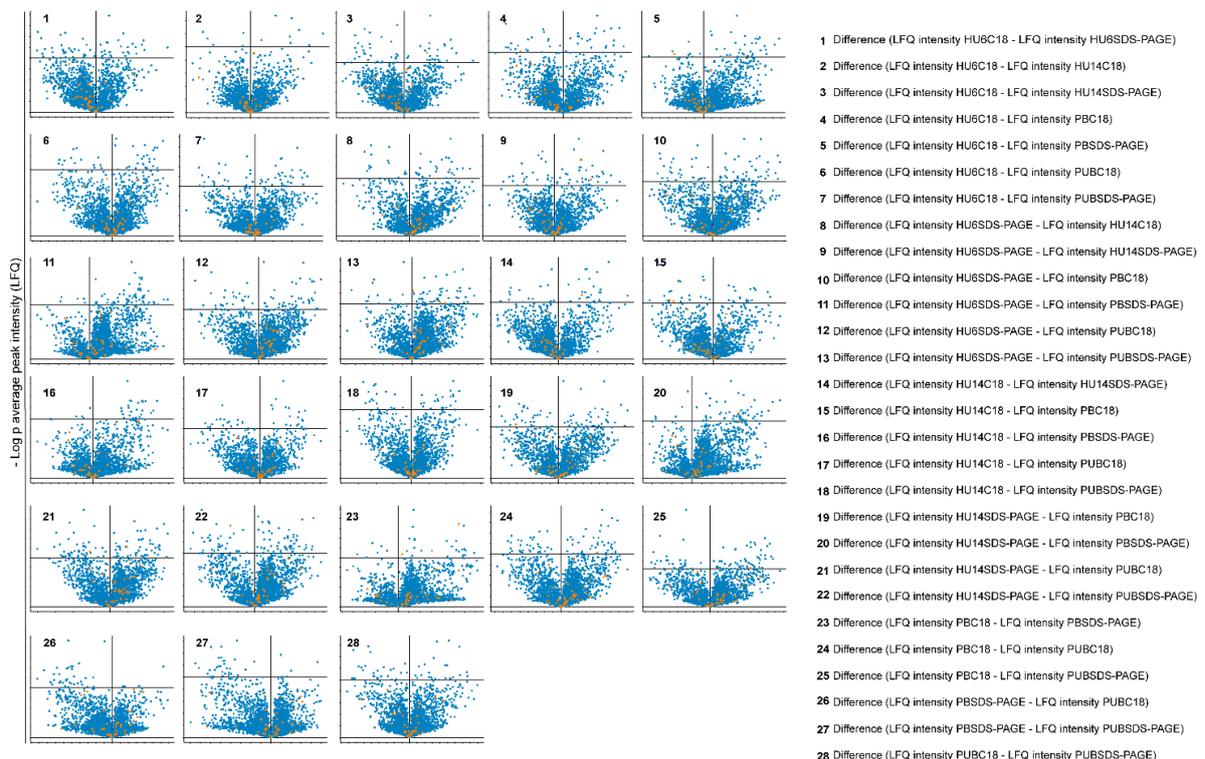


Figure 3.11. Volcano plot Analysis: Volcano plot showing significantly differentially abundant proteins in different 28 comparisons of the tested methods. The plots show the difference of LFQ intensities between method pairs (x-axis) versus the $-\log_{10}$ p-values (y-axis) for each protein. The plots were based on $n = 3597$ proteins, and the horizontal bar in each plot represents $-\text{Log } P$ value 2.5.

3.3.6 Human protein atlas-based tissue annotation.

To attempt an analysis of the likely tissue origins of the plasma proteins identified in our workflows, I searched for tissue specificity of our total 3,597 commonly identified high confidence proteins within the Human Protein Atlas (HPA) (<https://www.proteinatlas.org/>) (29-31). I provided a list of 51 organs/tissues in which our 3,597 HPA curated proteins can be found in supplementary data presented in the full paper in the appendix. To obtain a global overview of the origin of our protein list, by specificity to organ systems, the 3,597 proteins were collapsed into 12 categories, which included the 11 main organ systems of the human body (Figure 3.12A). Next, I sought to map the brain-specific proteins using the human brain protein atlas. The analysis mapped an average of 1430 detected proteins in all the ten brain sub-regions, while a specific number of mapped proteins are reported in Figure 3.12B. The classification of the elevated regional level of proteins showed an average of 14 proteins while the minimum was found in the thalamus (1) and midbrain (2) and maximum in pons and medulla (39) and basal ganglia (28). The differential high abundance protein identification in our plasma proteins study reflects its high potential for the use of brain-specific biomarkers in altered physiological conditions.

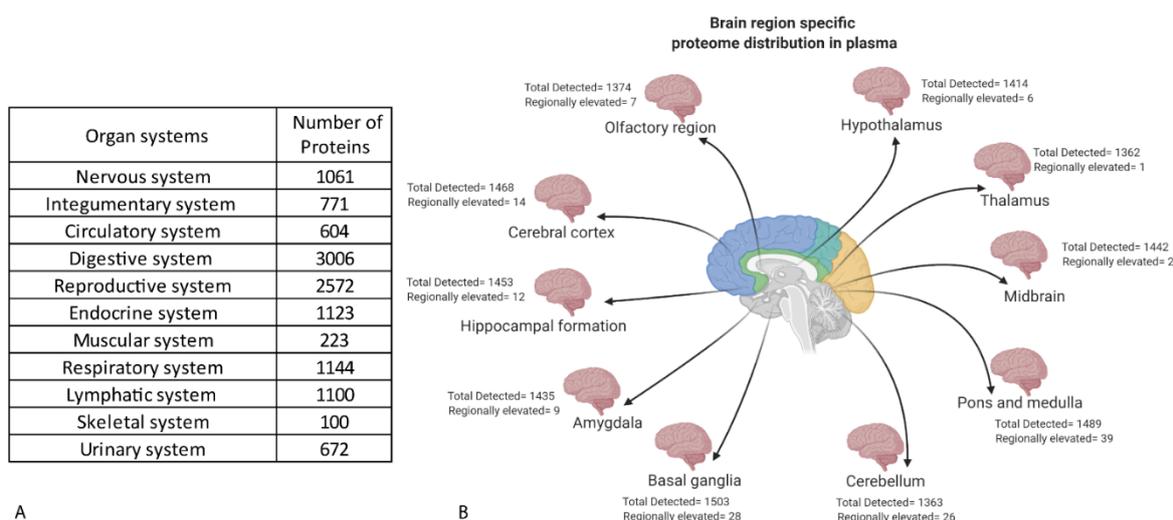


Figure 3.12. A. Table is representing the proteins identified in the plasma proteome, representative of the 11 central organ systems; nervous system (spinal brain cord, sensory organs): digestive system (mouth, oral mucosa, salivary gland, esophagus, intestine, stomach, colon, duodenum, gall bladder): muscular system (heart, skeletal, smooth): skeletal system (bone, cartilage, ligaments, tendons): integumentary system (skin, hair, nails, breasts, soft tissue, adipose tissue): reproductive system (ovaries, testis, vagina, uterus, cervix, endometrium, epididymis, fallopian tube, prostate, seminal vesicle): urinary system (kidneys, bladder): circulatory system (liver, blood, blood vessels, bone marrow): endocrine system (pituitary, thyroid, parathyroid, adrenal, pancreas): lymphatic system (lymph nodes, spleen, thymus, appendix, tonsil): respiratory system (lungs, nasopharynx). Each of the proteins in our combined dataset (3,597 proteins) was manually checked against the Human Protein Atlas (HPA) (<https://www.proteinatlas.org/>) to determine which proteins are most abundantly expressed in 51 different organs. B. Similarly, the whole proteins were mapped to identify the brain-specific sub-organs using the Human brain atlas (<https://www.proteinatlas.org/humanproteome/brain>). The regionally elevated proteins identified were reported separately, while all the proteins specific to relevant organs were also reported.

3.4 Discussion

The eight methods of plasma high abundance protein depletion and low abundance protein enrichment evaluated in this study allow considerable enhancement of typical plasma proteome coverage, with a total number of 4413 proteins identified with high confidence using LC-MS/MS (data-dependent analysis) and database searching with MaxQuant and TransProteomic Pipeline. Plasma is arguably one of the most challenging sample types for this kind of analysis. Nevertheless, I have demonstrated that good proteome coverage can be achieved even on this quantitatively and qualitatively complex proteome if appropriate sample processing approaches are applied. Furthermore, our results compare well with other reported work aimed to maximize plasma proteome coverage^{236,237}.

The human genome project has identified approximately 20000-25000 human genes (<https://www.genome.gov/human-genome-project>), while the human proteome draft article reported the identification of 17,294 proteins in total using mass spectrometry²³⁸. In our analysis, I identified a total of 4413 confidently identified proteins across all workflows, representing 3766, 3792, 3470 and 3682 proteins in the HU6, HU14, PB, and PUB workflows, respectively. Therefore, our estimate of the human proteome coverage, as a percentage of the 17,294 proteins reported by Kim et al 2014 are as follows; total = 25.52%, HU6 = 21.78%, HU14 = 21.93%, PB = 20.06%, and PUB = 21.29%. The average coverage I identified is 21.27% with standard error (SEM) = 0.42%.

Of the 4385 total proteins identified, 3064 (ca. 70%) overlap across methods, and excellent correlations are observed for all within-method technical replicates, and even reasonable correlations for across-method technical replicates were obtained. Consequently, I observe that all eight methods performed well in facilitating the identification of high numbers of proteins, and all achieve comparable protein identifications to the highest of recent reports detailed in^{223-235,239,240} Table 3.2. However, a recent literature review observed that of ~180 plasma proteomics studies (conducted during 2005–2017), only 50% reported identifying 500 of the most abundant plasma proteins, following reanalysis of data sets using PeptideAtlas²⁴¹. This is likely a reflection of the enthusiasm for minimal sample processing since most approaches perform at most a single-step fractionation, usually focused on the removal of one or more of the highest abundance proteins.

Table 3.2. Mass-Spectrometry-Based Plasma Proteomic Studies Allowing Comparison of Methodology, Run Time, and Numbers of Proteins Identified.

S.No.	Title of the study	Depletion strategy	Number of fractions/samples × run time for each fraction/sample (min.)	Total run time (min.)	Total number of proteins	PRIDE accession
1	Affinity Capture Enrichment versus Affinity Depletion: A Comparison of Strategies for Increasing Coverage of Low-Abundant Human Plasma Proteins.	Affinity-based probes based on agarose-immobilized benzamidine (ABA), O-Phospho-L-Tyrosine (pTYR), 8-Amino-hexyl-cAMP (cAMP), or 8-Amino-hexyl-ATP (ATP) HU14 Agilent technologies	100 × 76	7600	Total proteins=1165 HU14= 422; Proteominer=590; ABA=598; ATP=449; cAMP=436; pTYR=404	PXD020727
2	In-depth plasma proteomics reveals an increase in circulating PD-1 during anti-PD-1 immunotherapy in patients with metastatic cutaneous melanoma	HU14 Agilent technologies	72 × 50	3600	1917	PXD017201
3	Blood plasma proteomic modulation induced by olanzapine and risperidone in schizophrenia patients	HU14 Agilent technologies	Patients group A 14 × 39 Patients group B 40 × 39	Patients group A= 546 Patients group B= 1560	Patients group A= 319 Patients group B= 251	PXD015049
4	Age-Dependent Changes in the Plasma Proteome of Healthy Adults.	ProteoMiner (Bio-Rad Laboratories, Hercules, CA, USA)	18 × 90	1620	1069	PXD016199

5	Glyco-CPLL: an integrated method for in-depth and comprehensive N-glycoproteome profiling of human plasma.	ProteoMiner column (Bio-Rad Laboratories, Hercules, CA, USA)	6 × 78	468	Healthy control= 567 Patients= 537	PXD016428
6	Identification of Potential Plasma Biomarkers for Abdominal Aortic Aneurysm Using Tandem Mass Tag Quantitative Proteomics	TOP 12 depletion kit (Pierce, Thermo Fisher)	27 × 90	2430	169	PXD011309
7	Proteomes of paired human cerebrospinal fluid and plasma: Relation to blood-brain barrier permeability in older adults.	HU14 Agilent technologies	120 × 150	18000	422	PXD009589
8	Plasma Proteome Profiles of Stable CAD Patients Stratified According to Total Apo C-III Levels.	Hybrid SPE(R)-Phospholipid, a RP-C4 cartridge for solid-phase extraction (SPE)	52 × 17	884	188	PXD005973
9	Temporal expression profiling of plasma proteins reveals oxidative stress in the early stages of Type 1 Diabetes progression.	HU14 Agilent technologies	21 × 120	2520	2235	PXD007884
10	Discovery of a Potential Plasma Protein Biomarker Panel for Acute-on-Chronic Liver Failure Induced by Hepatitis B Virus	HU14 Agilent technologies	14 × 90	1260	427	PXD007975
11	Proteomics reveals the effects of sustained weight loss on the human plasma proteome	HU6 Agilent technologies followed by ProteoPrep20 Plasma Immunodepletion Kit	52 × 45	2340	1294	PXD004242

12	Plasma Proteome Profiling Reveals Dynamics of inflammatory and Lipid Homeostasis Markers after Roux-En-Y Gastric Bypass Surgery	High Select Top14 Abundant Protein Depletion Midi Spin Column Multiple Affinity Removal Spin Cartridge Human 6	24× 45	1080	1928	PXD009348
13	Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury	IgY14 LC20 followed by Supermix LC10 column	30× 172	5160	4641	Not Submitted
14	My present study	1. Hu6 C18 2. Hu6 1D SDS PAGE 3. Hu14 C18 4. Hu14 1D SDS PAGE 5. PB C18 6. PB 1D SDS PAGE 7. PUB C18 8. PUB 1D SDS PAGE	24 × 60 (equal run time for all eight-methods combination tested)	1440	1. 3364 2. 3752 3. 3778 4. 3508 5. 3421 6. 3157 7. 3663 8. 3813	PXD02246

For initial fractionation, both high abundance protein removal (Hu6, Hu14) and low abundance protein enrichment (ProteoMiner beads) perform well in reducing the plasma protein dynamic concentration range. One advantage of high abundance protein removal over low abundance protein enrichment is that the antibody-based immunoaffinity chromatographic techniques (Hu6 and Hu14) are highly effective in the specific removal of the highest abundance proteins, with minimal non-specific loss of other proteins²⁰⁴. By comparison, the bead-based (ProteoMiner) method enriches the lower abundance proteins but also loses much of the proteome to the “unbound” fraction (*i.e.*, bead saturation), which risks the loss of quantitative data (unless both the bound and unbound fractions are analysed), thereby increasing time and complexity for discovery-based quantitative projects. Previous studies have also shown almost equal numbers of proteins in bound and unbound fractions, and a large overlap (91% identified proteins) was found between crude samples and depleted unbound fractions. Our data support these observations²⁴². I identified such a high number of bound proteins due to the extensive fractionation strategies I applied post-low-abundance protein enrichment. Most proteomics studies that have employed the ProteoMiner kit with no additional fractionation report identification of ~500–1000 proteins^{223,226,227,243}.

While analyzing the ProteoMiner unbound fractions (flow-through), I identified 3,682 total proteins in the flow-through, indicating that the fractionation facilitated the identification of increased numbers of proteins and demonstrated bead saturation of a vast number of proteins, which would be problematic for subsequent quantitative work. To overcome this quantification issue, it is possible that the proteins uniquely expressed in the bound fraction but not present in the unbound fraction could be compared in a relative quantification approach since these are not saturated. Then the saturated proteins could be quantified in the unbound fraction, where the depletion via bound fraction may even enhance quantitative differences in case vs control comparisons. However, this would add complexity, time and cost to the methodology since both fractions would need to be assayed.

In the second phase separations (C18 vs SDS PAGE), higher overall LFQ intensities, and in most cases also peak counts, were achieved in the Hu6 and Hu14 depleted samples, with the methods which included SDS PAGE fractionation. It is unclear why this should be, but the LC fractionation may dilute the peptide fractions, so peptide losses are likely to be more significant as the diluted peptides bind non-specifically to tube surfaces. By contrast, SDS PAGE does not dilute the sample and traps and concentrates proteins within gel bands until they are liberated

during in-gel tryptic digest. The sample volumes are minimal throughout these steps, providing fewer opportunities for non-specific binding. These problems with the C18 fractionation might likely be overcome using higher protein concentrations to start with and/or low protein binding tubes. Here I used a starting amount of 50µg of low abundance plasma proteins for all methods to directly compare the 8 workflows. The amount of protein used reflects a level of low abundance protein that can easily be recovered from a 50 -100µL aliquot of plasma. This is an important consideration for clinical and biobank samples, where plasma from population-based studies can be limited and demand a broad range of studies. Procedures which result in minimal sample loss are important in this context. Apart from minimizing sample volume required while maximizing data output, sample throughput is another vital consideration if clinical studies are completed in a reasonable timeframe. A further advantage of SDS PAGE is that it is inherently a higher throughput method than C18 LC because 10s – 100s of samples can efficiently be run in parallel during the 1hr gel run time and in a relatively cost-effective manner since the equipment for SDS PAGE is considerably less costly than HPLC equipment. The C18 approach allows only serial separation, thereby greatly increasing sample preparation time in direct proportion to the number of samples. Furthermore, equipment and running costs are considerably lower for SDS PAGE than for HPLC. Our observations are supported by other published work demonstrating a throughput advantage of gel-based separation vs LC fractionation²⁴⁴.

Identifying and quantifying over 10,000 proteins in samples like tissue and cell lines is possible without extensive prefractionation²⁴⁵. This is true because while cellular proteomes are complex in terms of numbers of proteins, their concentration range is much less so (likely 10^3 - 10^4)⁵³. By contrast, the concentration range of plasma proteins has been reported to be as high as 10^{13} ²³⁹, with just a handful of proteins representing >80% of the total protein content, while 1000s of proteins are represented within the remaining <20% fraction. Consequently, plasma fractionation is unavoidable to achieve protein coverage of the medium/low abundance proteins in plasma, particularly the lower abundance tissue-specific proteins that may reflect the disease-specific change. In 2015, Keshishian et al. demonstrated this point by fractionating plasma and reported the identification of ~5,000 proteins from 16 plasma samples, using high pH reversed-phase separation in combination with iTRAQ 4-plex labelling²³⁵. This work represents one of the most comprehensive plasma proteomics profiling projects to date, utilizing a 3 step procedure of fractionation, including; (1) depletion of the top 14 HAPs (IgY14

LC20 column), (2) removal of the next ~50-100 moderate abundant proteins (MAPs) using a Sigma-Aldrich Supermix LC10 column, and (3) tryptic digestion of the remaining proteins and fractionating by C18 HPLC²³⁵. This powerful approach was an excellent demonstration of the effectiveness of fractionation to yield high coverage of the plasma proteome. However, the methodology is also complex, expensive, time consuming and unlikely to be practical for many clinical projects, thereby restricting general utility and replication of results in other laboratories. Consequently, cost effective and easily accessible plasma proteomics analysis remains a challenge.

Here, the work demonstrates that more straightforward and faster approaches, accessible to most general biochemistry laboratories, can yield similarly high proteome coverage. Some examples of potential biomarkers which are usually present in plasma at ng/ml concentrations were identified in our data, including Superoxide dismutase (SOD2), Ribonuclease 4 (RAB4), 72 kDa type IV collagenase (MMP2), Alpha-1-antitrypsin (SERPIN) proteins. Among these are some proteins with a role in disease, such as RAB4, which protects neurons from degeneration in amyotrophic lateral sclerosis (ALS) by stimulating neurofilament formation and protecting hypothermia-induced degeneration in mouse embryonic cortical neurons²⁴⁶. Another protein identified in our dataset is matrix metalloproteinase 2 (MMP2), which is involved in remodelling vasculature, neural progenitor cell migration, and tumour invasion²⁴⁷. Some of the lowest low abundant plasma proteins that I identified include Macrophage colony-stimulating factor 1 (200 pg/mL), P-selectin (120 pg/mL), platelet-derived growth factor receptor β (3 pg/mL), and Leptin (4.7 ng/mL), some of which have also previously been reported²⁴⁸.

Our study shows that a variety of approaches can achieve the depth of plasma proteome coverage²⁴⁹. Two critical points appear to be (*a*) removal of the relatively few highest abundant proteins, which represent >60-80% of plasma total protein, and (*b*) fractionation of the remaining medium-low abundance protein fraction. Based on the 8 methods evaluated, I suggest the following: (*1*) with a view to subsequent quantification, high abundance protein removal is preferable to low abundance peptide enrichment to avoid the bead saturation problem; (*2*) for high throughput where analysis of high numbers of clinical samples is required (10s – 1000s), SDS PAGE is a faster approach than C18 as a secondary fractionation approach. Consequently, of the variety of methods evaluated here, the HU6-SDS PAGE and HU14-SDS PAGE approaches best satisfy all requirements, including narrowing the dynamic

concentration range and the resultant depth of proteome coverage, retaining quantitative characteristics of the sample and relatively simple, cost-effective and time-efficient workflow. There is no doubt that plasma proteomics can provide high confidence diagnostic tools for better stratification and monitoring of patient response to treatment^{212,213,235,249}.

3.5 Conclusion

The primary objective was to design a fractionation approach that would offer plasma proteome coverage and identify approximately 4,000 tissue specific proteins. The results demonstrated the reliability of fractionation techniques as well as the benefits of gel-based tests over costly and time-consuming chromatographic procedures. The results are relevant in plasma proteomics research and to the diagnosis of probable AD. In particular, human plasma is one of the most extensively utilised tissues in clinical analysis, and plasma-based biomarkers are employed for monitoring patient health status and/or response to medical therapy to prevent needless invasive biopsy. This is partly because of the complexity of the plasma proteome, especially its wide quantitative dynamic range, estimated to be between 9 and 13 orders of magnitude between the protein with the lowest abundance and the one with the greatest abundance in data-driven plasma proteomics. A key problem is to find methods that may accomplish depth of plasma proteome coverage while limiting the complexity of the sample workup and increasing the sample throughput. In this study, we have used the Agilent multiple affinity removal liquid chromatography (LC) column, the Agilent multiple affinity removal LC column, and ProteoMiner to deplete high-abundant plasma proteins or to increase the amount of low-abundant plasma proteins. We then used sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS PAGE) and C18 prefractionation techniques. We analysed the performance of each of these fractionation procedures in order to determine which method best meets the criteria for clinical sample analysis while also providing enough plasma proteome coverage and suitable sample yield. Comparing one-dimensional (1D) gel-based prefractionation to serial chromatographic separation, we demonstrate that it allows for parallel sample processing with no loss of proteome coverage, and that it significantly reduces analysis time, which is particularly important for large clinical projects. Additionally, we show that a range of techniques may produce comparable levels of plasma proteome coverage, providing for method selection flexibility depending on project-specific requirements.

Limitation:

A potential drawback of immunodepletion techniques is the elimination of nontargeted proteins by nonspecific interaction with depletion columns or trapped proteins. The bound fraction associated with immunodepletion columns comprised nontargeted proteins, the majority of which were detected at low levels using spectral counts. This observation is consistent with prior findings of IgY-14 columns capturing nontargeted proteins. Our results suggest that binding is repeatable for nontargeted proteins acquired at greater quantities. Along with continuously depleting targeted proteins and producing little depletion of nontargeted proteins, immunodepletion columns significantly increased detection of nontargeted proteins, often by roughly 4-fold, based on spectral counts.

Significance:

This work reveals that a variety of basic prefractionation procedures are capable of achieving the depth of coverage required for the plasma proteome. Additionally, the findings indicate the reliability of these pre-fractionation procedures and suggest that gel-based methodologies may be used in place of costly and time-consuming chromatographic separation, greatly shortening the time required for analysis. Additionally, the fact that a variety of methodologies can achieve comparable levels of proteome coverage provides flexibility in terms of project-specific requirements, such as whether qualitative or quantitative information is sought, the size of the project in terms of sample numbers, and the availability of specific laboratory resources. These factors are critical in the attempt to speed plasma proteomics research in order to enable rapid, reliable, and accurate diagnosis, population-based health screening, clinical research investigations, and other clinical activities.

Chapter 4

Deep proteome analysis of plasma reveals novel biomarkers of Alzheimer's disease: A longitudinal study

Manuscript

4.1 Introduction:

Alzheimer disease (AD) accounts for up to 70% of all dementia cases and is the most common cause of dementia. Ageing is the primary risk factor for AD; however, there is a poor understanding of the biological mechanisms by which the ageing process contributes to AD development in some individuals, while others progress to advanced age with relatively little AD neuropathology.

The pathogenesis of AD is now recognized to be multifactorial, with dysregulation of various cellular and molecular processes contributing to the disease process, including synaptic damage, mitochondrial dysfunction, and oxidative stress^{18,100,250-253}. While advancing age is the single greatest risk factor for AD²⁵⁴, other factors such as *APOEε4* allele¹³, comorbidities such as vascular disease²⁵⁵ and lifestyle factors such as head injuries^{256,257} all contribute to the level of AD risk. Early AD manifests clinically as mild cognitive impairment (MCI)¹⁵⁴, particularly in the case of amnesic MCI, although a clinical diagnosis of MCI often stays stable or even reverts to normal and does not always progress dementia^{258,259}. By the time AD manifests as dementia, the level of brain pathology may be impossible to revert since substantial neuronal cell death has occurred. Identifying biomarkers of transition from normal to MCI (if not earlier) might provide a window of opportunity for prevention trials that focus on ameliorating symptoms before neurodegeneration progresses to clinically identifiable symptoms. Several neuroimaging and CSF based biomarkers for diagnostic evaluation of dementia have recently been recommended by an international consensus group¹⁹⁹. The major limitations with CSF and neuroimaging biomarkers are that they are not likely to be widely adopted for routine use or population screening due to their invasive nature, high cost, limited availability and requirement of high-level technical skill and training to implement. By contrast, blood is a relatively easy fluid to collect, and venepuncture is a routine and commonly performed procedure for clinical and research purposes.

Mass spectrometry-based methods represent the only unbiased approach for discovery focused proteome analysis. They are rapid, sensitive, can provide both qualitative and quantitative information, and for the study of proteins, can also provide information about post-translational modifications and protein interactions. The main obstacle has been identifying methods of narrowing the extreme dynamic range of the plasma proteome while maintaining sufficient methodological simplicity to apply to moderately sized clinical studies. Recent advances in

plasma proteomics have identified promising approaches to achieve the depth of plasma proteome coverage using prefractionation methods^{38,249,260}. In the current study, I used a two-step plasma fractionation approach; HU14 removal of high abundance plasma proteins, followed by parallel 1D SDS/PAGE of the low abundance protein fractions, based on our previously published method³⁸. This workflow facilitated extended plasma proteome coverage unbiased, allowing identification of biologically meaningful longitudinal and cross-sectional proteome changes in individuals progressing through stages of cognitive impairment over the decade in which greater risk commences (65-75+ years). A set of potential protein biomarkers might facilitate the development of precise tests for detecting the disease at the early stages. Furthermore, these markers may help identify unexpected biological pathways and new potential therapeutic targets for future development.

4.2 Materials and Methods:

4.2.1 Experimental procedures

Cohort, plasma and experimental proteomics design

Plasma samples were obtained from the Sydney Memory and Ageing Study (MAS) from participants aged 70-90 years²⁶¹. The baseline sample was collected (Wave 1) between September 2005 and November 2007, at which time all participants were cognitively normal (n = 33). Participants were followed up for six years (Wave 4), with 11 participants remaining normal and the remainder progressing to MCI and AD (n=11 each). The diagnosis was by consensus and met the NIA-AA criteria for MCI and AD, respectively (Table 4.1). Detailed inclusion and exclusion criteria for the MAS cohort was previously published²⁶¹. I selected only individuals with aMCI (amnestic MCI) for this study, as this subtype is generally related to subsequent progression to Alzheimer's dementia^{262,263}. Additionally, I analysed samples from individuals with a clinical diagnosis of dementia, probable AD, henceforth to be abbreviated as AD.

Blood was collected into EDTA containing tubes, centrifuged (2000g, 20min, 4°C), and the plasma transferred into clean 1.5mL polypropylene tubes. To minimize freeze-thaw cycles, plasma aliquots were prepared (50, 250 and 500 µl) and stored at -80°C until required. The

UNSW Human Research Ethics Committee approved a protocol for blood collection (MAS ethics number; HC14327).

Proteomics profiling was performed on 33 humans (66 total) plasma samples from wave 1 (baseline) and wave 4 (6 years follow up), in the following three groups: (1) individuals cognitively normal at wave 1 denoted as CTRLW1, who remained normal at wave 4 denoted as CTRLW4, (2) individuals cognitively normal at wave 1 denoted as MCIW1 who progressed to MCI at wave 4 denoted as MCIW4, (3) individuals cognitively normal at wave 1 denoted as ADW1 who progressed to dementia, probable AD at wave 4 denoted as ADW4.

4.2.2 Depletion of high abundant proteins using Human 14 (HU14) immunoaffinity-based columns:

The protocol followed for plasma high abundance protein removal, and fractionation of the low abundance proteins was adapted from Chapter 3³⁸. The approach involved depletion of the top 14 high abundance proteins (albumin, IgG, antitrypsin, IgA, transferrin, haptoglobin, fibrinogen, α -2-macroglobulin, α -1-acid glycoprotein, IgM, apolipoprotein AI, apolipoprotein AII, complement C3, and transthyretin) using an Hu14 column (4.6 x 100 mm, Agilent California, United States), followed by SDS/PAGE fractionation of the low abundance protein fraction.

4.2.3 Fractionation of low abundance proteins using 1D-SDS PAGE, tryptic digest and LCMSMS

Equal amounts of total protein (50 μ g) from the Hu14 depleted plasma were filtered using Amicon ultra 3kDa centrifugal filter units (MERCK, New Jersey, USA), dried in speed vac (ThermoFisher, Massachusetts, USA) and diluted to a final volume of 20 μ L by adding 5 μ L LDS sample buffer Invitrogen NuPAGE (ThermoFisher, Massachusetts, USA), 2 μ L reducing agent Invitrogen NuPAGE (ThermoFisher, Massachusetts, USA), and 13 μ L deionized water (MilliQ). Samples were then briefly heated (10 minutes, 70°C), followed by electrophoresis; 1D SDS/PAGE using Invitrogen NuPAGE 4-12% gradient Bis-Tris midi gels (ThermoFisher Scientific, Massachusetts, USA) and 1x Invitrogen MES running buffer according to the manufacturer's instructions (ThermoFisher Scientific, Massachusetts, USA, USA) followed by

colloidal coomassie G250 staining²⁶⁴ (Figure 4.1). After destaining, the separated protein lanes were cut evenly with a 24-lane blade (Gel Company Inc., CA), and the gel slices were collected into ten vials for destaining, in-gel trypsin digestion and label-free LCMSMS, following the approach taken in our previously published work³⁸.

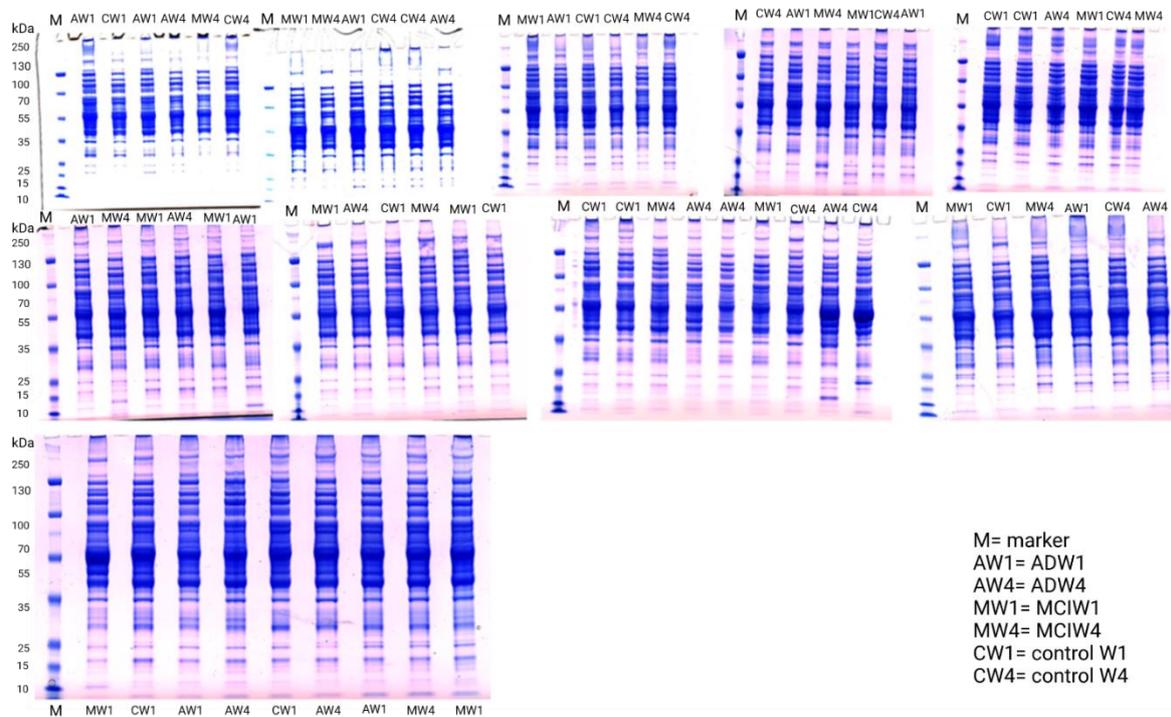


Figure 4.1. Representative image of NuPAGE LDS gel profile of and depleted plasma containing low abundant plasma proteins (LAP) from HU14 from all the individuals. Each gel lane contained an equal loading of total protein (50 μ g total proteins were loaded per gel lane).

4.2.4 Computational Analysis

Computational analysis of the raw files was performed for protein identification and quantification. The consistency of protein expression change was determined using two label-free quantification approaches, peak area integration and spectral counting. Protein identification, peak area integration and fold-change calculation were performed using ProteomeDiscoverer v2.4 software (Thermo Fisher Scientific, Waltham MA), in conjunction with three search engines (Mascot, Sequest, and Amanda). Protein identification followed by spectral counting and fold-change determination was carried out using a combination of Mascot search engine and Scaffold Q+ software v 4.11.0 (Proteome Software, Portland, OR). A minimum of ≥ 2 unique peptides per protein were required for protein identification and quantitation on all data analysis software. The UniProt *Homo sapiens* (human) database was combined with reversed decoy database to determine FDR by all search engines for MS and

MS/MS spectral mapping. Mass tolerance for matching peaks to theoretical ion series was five ppm. False discovery rate (FDR) was set to <1% to ensure only high-confidence protein identifications. Enzyme specificity was set to trypsin, with a maximum of two missed cleavages. Searches included variable modifications of protein N-terminal acetylation, methionine oxidation, and fixed modification of carbamidomethylation of cysteines. All the parameters were kept similar in both search engines. To select only those proteins with robust expression change between groups, I used the following inclusion criteria: only those proteins quantified in >6 individuals, proteins identified with a minimum of two peptides per protein, the consistent direction of protein fold change across two bioinformatics platforms with orthogonal quantification approaches (peak area ratio with PD2.4 and spectral counting with Scaffold) with a fold change of at least 20% (≤ 0.08 and ≥ 1.2) in preferably both search engines but at least one. These orthogonal approaches have specific advantages and disadvantages^{265,266}, so I reasoned that the most reliable changes should be consistent across platforms.

4.2.5 Bioinformatics Analyses

I used RStudio version 1.2.5033 and R version 3.6.3 for most post data processing analyses, including heatmap and volcano plots. Venny 2.1 was used to plot Venn diagrams²⁶⁷. I performed gene ontology (GO) term enrichment analysis using differentially expressed proteins (DEPs) to compare biological processes and pathways affected in normal ageing, MCI and AD, using STRING (version 11.5). This kind of analysis uses GO terms to classify proteins into particular roles or functions (i.e., biological processes, cellular components, molecular function and KEGG & Reactome pathways). From this kind of sorting, I can identify numbers of proteins that subserve specific functions (i.e., “observed gene count” within the STRING output). Information about the level of enrichment of functional categories is also provided by comparison with a background set of proteins (I used the default whole human genome available within STRING for the analyses presented here), which allows an estimation of the enrichment score (strength) and level of statistical significance (FDR). Together the observed gene counts and enrichment strength values give an idea of which functional categories are represented by (a) the most significant number of proteins and (b) are most enriched relative to the background set. Both observations help identify functional categories that are associated with the disease. However, it should be noted that (1) most proteins are pleiotropic and may be listed within several functional groups, and (2) the GO term lists are a manually curated artificial construct and include some very broad terms which may capture many proteins (e.g.,

cellular process, biological regulation, binding, and others), but which are minimally informative from a specific function perspective. For this reason, observed gene count and enrichment strength values generally vary in an approximately reciprocal manner and therefore should be used together to identify biological/disease relevance functions. It is likely that categories of the greatest relevance will be those with a moderate score for both observed genes count and enrichment strength, rather than those that fall at the extremes of either value.

4.3 Results:

4.3.1 Overview of proteomics study populations:

The main objective of this study was to discover detailed plasma biomarker profiles reflecting normal ageing, mild cognitive impairment (MCI) and dementia, probable Alzheimer's disease (AD). Participant demographics are shown in Table 4.1.

Since each plasma sample consisted of ten fractions, a total of 660 LC-MSMS runs were performed to maximize plasma proteome coverage of low abundant proteins. In total, I identified 1,578 proteins (false discovery rate <1%) with 32,469 total peptides using the Proteome Discoverer 2.4 search engine. Data analysis was performed on 2 different search engines, i.e., Proteome Discoverer 2.4 and Scaffold Q+ software v 4.11.0. I performed analyses in seven different combinations, including both longitudinal and cross-sectional analyses. Longitudinal analyses included: **1.** Ageing while maintaining normal cognition (CTRLW4/CTRLW1) **2.** MCI (MCIW4/MCIW1); **3.** AD (ADW4/ADW1) and cross-sectional analyses included; **4.** MCI vs age-matched controls (MCIW4/CTRLW4); **5.** Incipient AD vs age-matched controls (ADW4/CTRLW4); **6.** MCI vs AD individuals (ADW4/MCIW4); and **7.** preclinical AD vs age-matched controls (ADW1/CTRLW1). The longitudinal analyses provide insight into changes that occur in normal ageing over 6 years and a progression from clinically normal to MCI or dementia, probable AD over 6 years of baseline to follow-up. These longitudinal analyses allow comparison of ageing while retaining clinically normal cognition and ageing with progression to cognitive disease and dementia, suggesting proteins and pathways which are disrupted in the development of disease/disorder. By contrast, the cross-sectional analyses compare incipient MCI or dementia and probable AD to cognitively normal age-matched controls, which may identify potential disease biomarkers.

Table 4.1: Details of the participant demographics which were included in our present study.

Total participants	Normal ageing	MCI	AD	Kruskal-Wallis statistic	Kruskal-Wallis P value
Total participants in each wave	11	11	11	NA	NA
Wave 1 age in years mean±SD (CV%)	76.89±3.39 (4.41%)	78.46±5.5 (7.01%)	80.62±4.72 (5.85%)	4.32	0.11
Wave 4 age in years mean±SD (CV%)	82.95±3.35 (4.04%)	84.37±5.59 (6.63%)	86.51±4.81 (5.56%)	4.38	0.11
Education (years) at wave 1	10.84±4.05 (37.44%)	10.77±3.92 (36.39%)	10.48±2.53 (24.16%)	0.04	0.97
Length of follow up (years)	5.77	5.95	5.75	NA	NA
Clinical diagnosis at W1	Normal	Normal	Normal	NA	NA
Clinical diagnosis at W4	Normal	amdMCI	Dementia, probable AD	NA	NA
W1 APOE status	E3/3	E3/3	E3/3	NA	NA
MMSE at W1 mean±SD (CV%)	29.36±1.50 (5.11%)	28.27±1.55 (5.50%)	28.36±1.80 (6.36)	5.61	0.06
MMSE at W4 mean±SD (CV%)	29.55±0.93 (3.16%)	28.18±1.40 (4.97%)	23.64±4.05 (17.16%)	20.92	0.00
Total WMH volume W1 mean±SD (CV%)	16032±19357 (120.7%)	9274±4112 (44.34%)	12612±13117 (104%)	0.28	0.87
Total WMH volume W4 mean±SD (CV%)	21407±14975 (69.96%)	17145±7380 (43.05%)	33005±20872 (63.24%)	3.40	0.18
BMI (median) mean±SD (CV%)	26.91±4.98 (18.54%)	28.73±6.05 (21.06%)	26.82±2.89 (10.78%)	1.01	0.60
Cholesterol (mmol/L)	5.08±0.98	4.41±0.75	4.79±0.72	3.41	0.18

mean±SD (CV%)	(19.32%)	(17.11%)	(15.09%)		
Triglyceride (mmol/L) mean±SD (CV%)	1.36±1.36 (99.76%)	1.26±0.71 (56.88%)	0.90±0.35 (39.75%)	1.47	0.47
HDL-Chol (mmol/L) mean±SD (CV%)	1.40±0.40 (28.99%)	1.20±0.35 (29.81%)	1.34±0.32 (24.05%)	2.16	0.33
LDL-Chol (mmol/L) mean±SD (CV%)	3.16±0.80 (25.46%)	2.64±0.66 (25.04%)	3.02±0.67 (22.31%)	2.71	0.25
Glucose (mmol/L) mean±SD (CV%)	6.58±2.76 (42.05%)	6.31±1.13 (18%)	5.80±0.58 (10.11%)	1.43	0.48
Urate (mmol/L) mean±SD (CV%)	0.32±0.07 (24.5%)	0.35±0.04 (13%)	0.35±0.08 (24.08%)	1.40	0.49
Vitamin A (umol/L) mean±SD (CV%)	2.80±0.48 (17.42%)	3.51±0.21 (20.10%)	3.15±1.27 (40.54%)	5.81	0.05
Vitamin E (umol/L) mean±SD (CV%)	44.13±33.77 (76.53%)	31.89±6.40 (20.08%)	31.74±9.14 (28.22%)	2.18	0.33
Carotene (umol/L) mean±SD (CV%)	1.03±0.76 (73.07%)	0.63±0.39 (62.94%)	0.79±0.52 (66.38%)	1.53	0.46

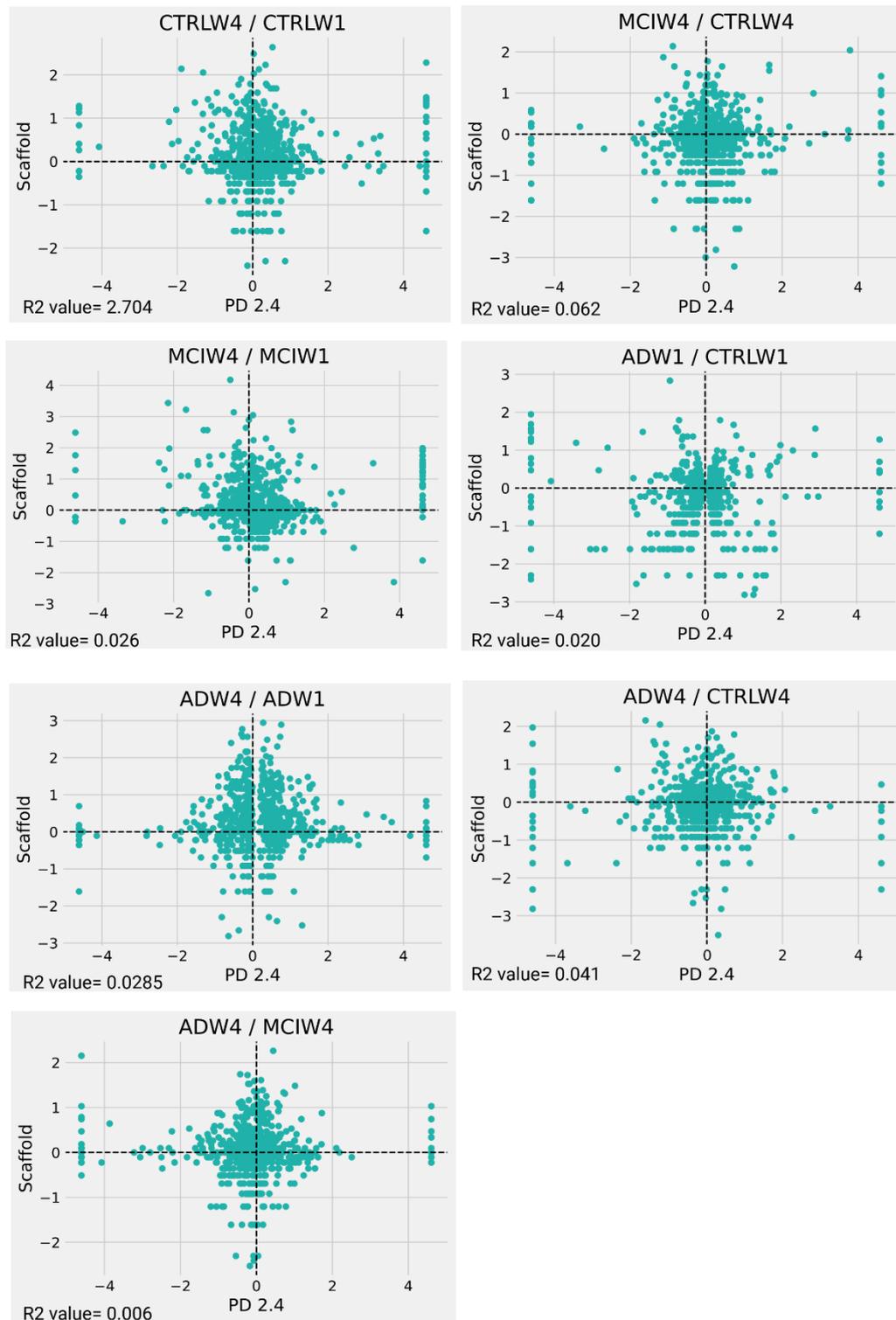


Figure 4.2i. Scatter plot and regression analysis of abundance ratio of DEPs in both PD2.4 and scaffold; CTRLW4/CTRLW1, MCIW4/MCIW1, ADW4/ADW1, ADW4/MCIW4, MCIW4/CTRLW4, ADW1/CTRLW1, ADW4/CTRLW4.

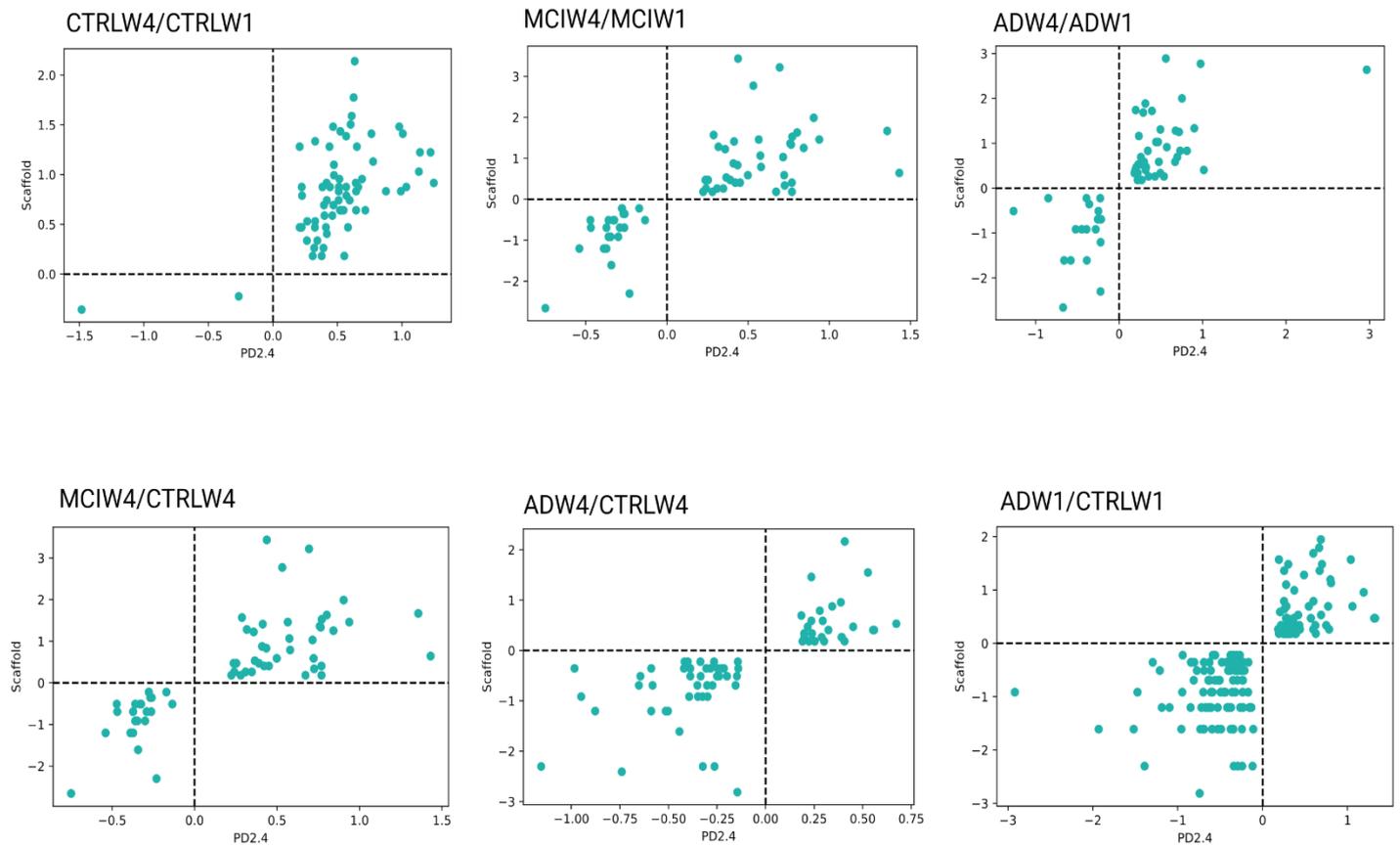


Figure 4.2iii: Density plot and regression analysis were plotted between all 6 comparisons of normal control, MCI, and AD in longitudinal and cross-sectional comparisons. Each dot represents the abundance ratio of each protein, and the colour shows the dot density. **S2iii.** Scatter plots were plotted using only DEPs from each comparison. **A.** 71 DEPs from CTRLW4/CTRLW1 (Table 4.3) **B.** 66 DEPs from MCIW4/MCIW1 (Table 4.5) **C.** 60 DEPs from ADW4/ADW1 (Table 4.4), **D.** 89 DEPs from MCIW4/CTRLW4 DEPs (Table 4.8), **E.** 70 DEPs from ADW4/CTRLW4 (Table 4.7), **F.** 160 DEPs from ADW1/CTRLW1 (Table 4.11)

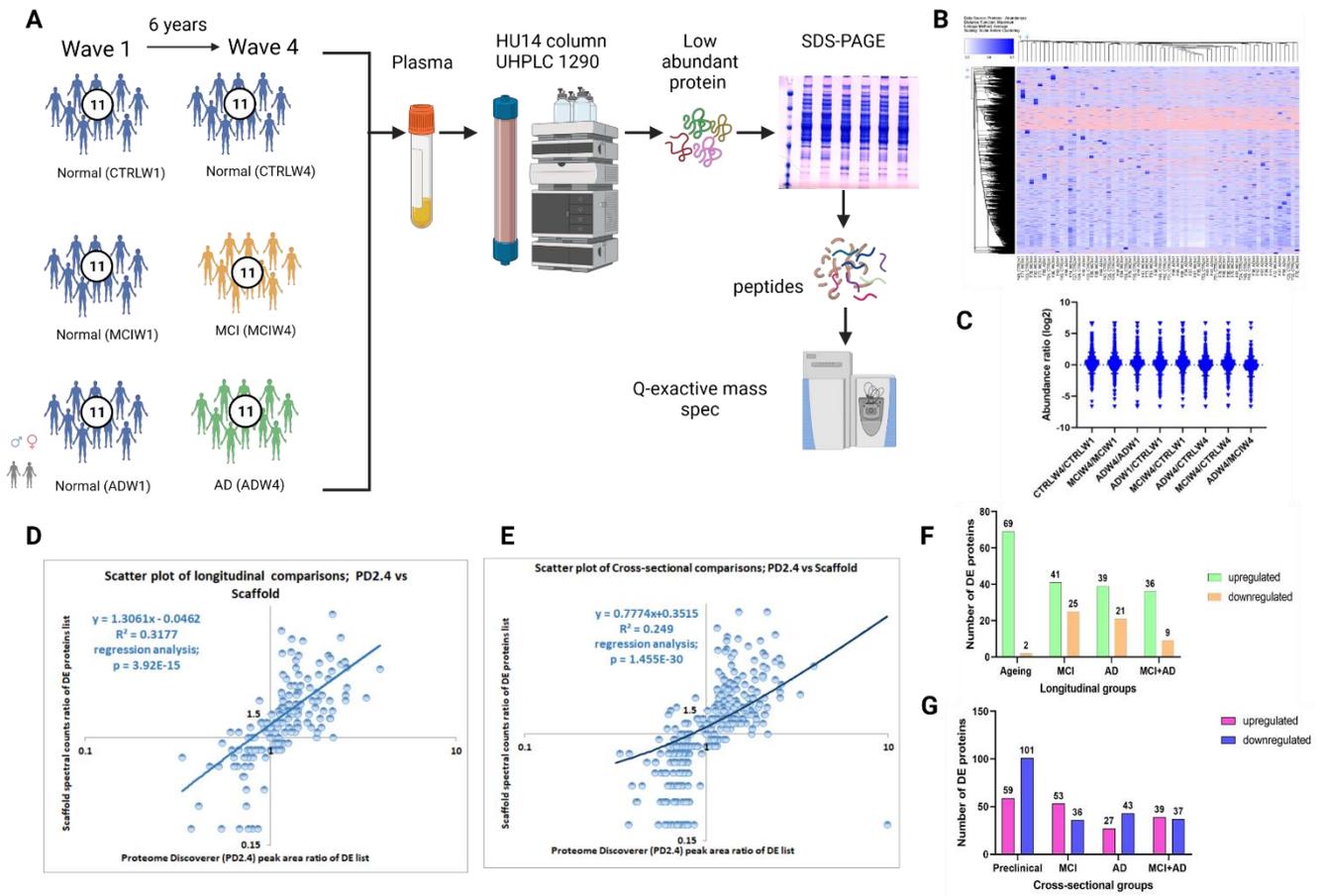


Figure 4.3i: Proteome profiling and comparison of normal ageing, MCI, and AD in longitudinal and cross-sectional cohorts.

Overview of the study population and schematic proteomic workflow. The plasma of two waves comprising ageing, MCI and AD subjects was analysed. The total number of subjects per group is depicted. Blue subjects represent normal individuals which aged normally in wave 4. Whereas orange and green depicting which progressed to MCI and AD from normal individuals in wave 4, respectively

Hierarchical cluster analysis and heat map for 1,578 total proteins identified in 66 individual samples (output from ProteomeDiscoverer 2.4 software).

Scatter dot plot analysis using abundance ratio of all 7 comparisons used in this study. Horizontal lines show the mean and the error bars \pm SD.

D and E. Scatter plots and regression analysis were plotted using the final list of DEPs used in both longitudinal and cross sectional comparisons.

F. Global analyses of proteomic changes in longitudinal groups. Bar graph showing the total number of proteins upregulated and downregulated in 1. Normal ageing (CTRLW4/CTRLW1), 2. MCI (MCIW4/MCIW1), 3. AD (ADW4/ADW1) and 4. MCI+AD (ADW4/MCIW4).

G. Global analyses of proteomic changes in cross-sectional analysis groups. Bar graph showing the total number of proteins upregulated and downregulated in 1. Preclinical AD (ADW1/CTRLW1), 2. MCI (MCIW4/CTRLW4), 3. AD (ADW4/CTRLW4) and 4. MCI+AD (ADW4/MCIW4).

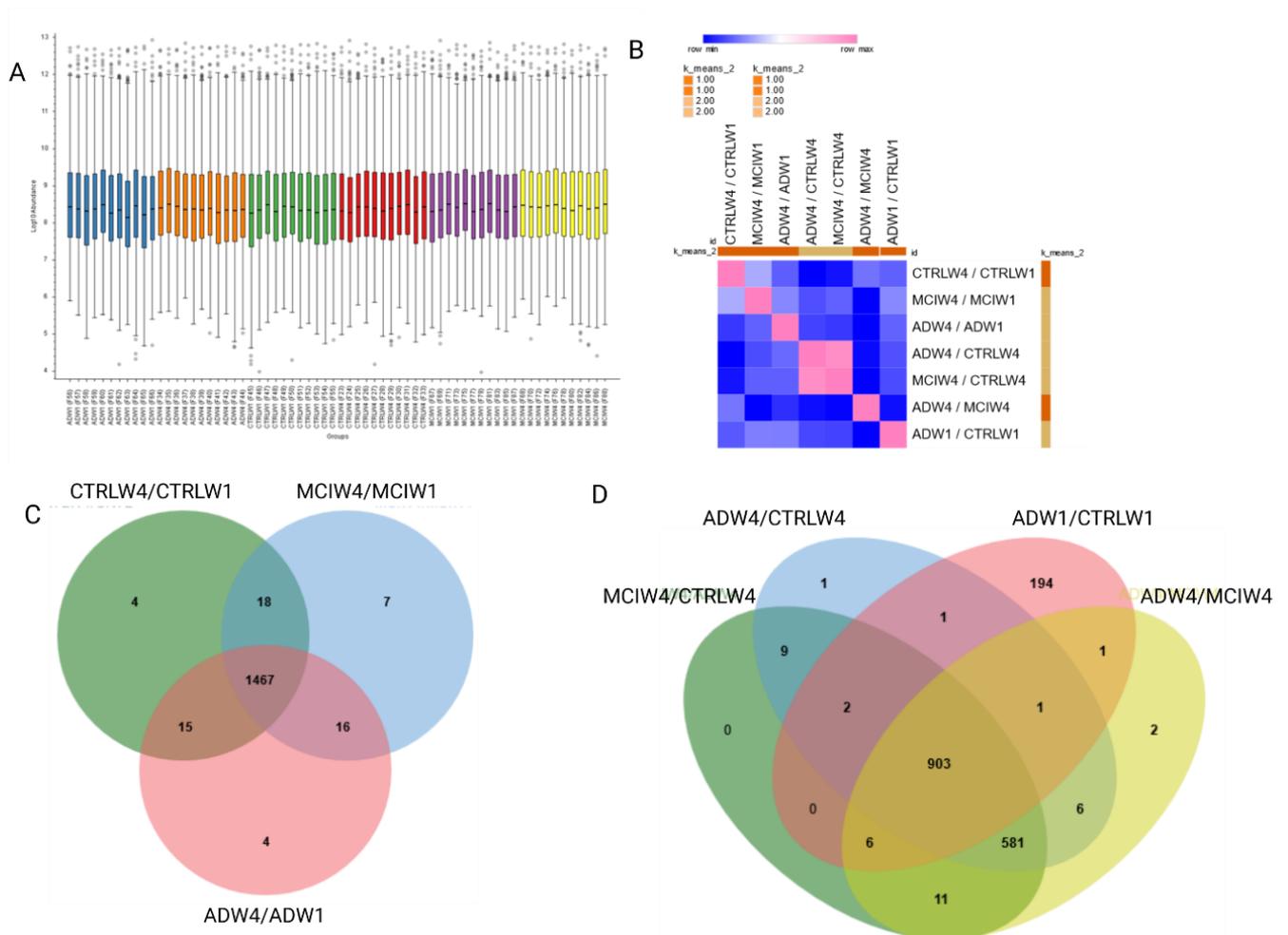


Figure 4.3ii:

A. Box and Whisker plots of abundance values of all 66 individual samples. The line within the box denotes the median value, and the upper and lower ranges of the box indicate the 5 and 95 percentiles of the abundance values, respectively (output from ProteomeDiscoverer 2.4 software).

B. The similarity matrix and heat map were constructed using the Pearson correlation values of the 7 comparisons, clustered based on the k mean algorithm.

C. Venn diagrams depicting the total number of proteins identified in longitudinal comparison. I identified 1467 proteins in all three longitudinal groups, i.e., normal ageing, MCI and AD.

D. Venn diagrams depicting the total number of proteins identified in cross-sectional comparisons. I identified 903 proteins in all four cross-sectional groups, i.e., preclinical AD, MCI, AD and MCI+AD.

The plasma proteomes of 33 individuals (11 individuals in each category; normal control, MCI, and AD) are compared by hierarchical clustering analysis (HCA) (Figure 4.3iB), abundance ratios (Figure 4.3iC) and box and whisker plot (Figure 4.3iiA), showing very similar distribution patterns overall. This is expected since most identified proteins' expression is unaltered between samples, even in disease. Similarity matrix analyses (Figure 4.3iiB) show a close association of protein expression data between the following group ratios: CTRLW4/CTRLW1, MCIW4/MCIW1, ADW4/ADW1 and ADW4/CTRLW4, MCIW4/CTRLW4. The two orthogonal methods of identifying differentially expressed proteins were compared using scatter plots and regression analyses (Figure 4.3iD and 1E),

showing significant regression between the two quantitative approaches (scaffold spectral counting and PD2.4 peak area integration). Bar graphs of the total number of proteins up and downregulated in longitudinal and cross-sectional comparisons are shown in Figures 4.3iF and 1G. In all longitudinal comparison groups, more proteins are upregulated than downregulated; this difference is particularly pronounced in the normal ageing group, with 69 upregulated and only 2 downregulated proteins (Figure 4.3iF). The numbers of proteins up and downregulated with age (over the 6 years of the longitudinal analysis) were similar in MCI and AD (Figure 4.3iF). In the cross-sectional comparison groups, the number of up and down-regulated proteins varies across groups (Figure 4.3iG), with MCI and AD having similar total numbers of DEPs. Interestingly, the preclinical AD group (Figure 4.3iG) had the greatest number of total DEPs, and also the more significant number of upregulated (59) and downregulated (101) proteins than either the incident clinical MCI or AD group.

Table 4.2: This summary table contains the final list of differentially expressed proteins (DEPs) in all the longitudinal and cross-sectional comparisons analysed. This list contains DEPs those quantified in >6 individuals, proteins identified with a minimum of two peptides/protein, the consistent direction of protein fold change across two bioinformatics platforms with orthogonal quantification approaches (peak area ratio with PD2.4 and spectral counting with Scaffold) with a fold change of at least 20% (≤ 0.08 and ≥ 1.2) in both search engines.

Comparisons analysed	Protein Gene Symbol
Age-related changes observed across all <u>longitudinal analysis</u> groups see Table 4.3 for details of fold change per group and p-value	Total protein number = 71 (69 Upregulated, 2 Downregulated) TPM4, CLIC1, ARHGDIB, YWHAZ, PAFAH1B2, TPI1, YWHAH, PGLS, ARPC3, PKM, PSMA4, ARHGDIA, GSTO1, GPI, ARPC2, YWHAB, YWHAE, SH3BGRL2, CFL1, NME2, LGALSL, ARPC1B, MAPRE1, PNP, TLN1, SERPINB1, ANXA5, GSTP1, PSME2, GAPDH, PSMB8, WDR1, ARPC4, ACTN1, PGAM1, FERMT3, PEBP4, MSN, ABHD14B, EIF5A, S100A9, TIMP1, CNN2, CLIC4, CMPK1, PARK7, LDHA, PPIB, FLNA, VCL, CALR, PPIA, PSMA5, YWHAQ, ARPC5, IGFBP2, RAB11A, ENO1, PSMA2, PGK1, ACTR3, LDHB, BIN2, OAF, CAP1, ILK, PRDX6, S100A4, TAGLN2, NCAM1, SELENBP1
MCI specific changes in <u>longitudinal analysis</u> (see Table 4.5 for details of fold change per group and p-value)	Total protein number = 66 (41 Upregulated, 25 Downregulated) RHEB, SSBP1, NUTF2, KRT35, C19orf10, UBE2V1, PSMB5, PLEK, COL1A1, PPBP, ITLN1, KRT86, HSPA4, HIST1H4A, IGFBP6, MB, CAPZA2, TNC, QDPR, PPP1R7, ARRB1, VCP, ABI3BP, RARRES2, ZYX, RNH1, APOD, ASGR2, RAB27B, S100A4, GLO1, CYCS, ADAMTSL4, COL5A1, GLIPR2, DDT, SERPINA1, LTF, GP1BA, HSP90B1, CDH2, IDH1, IGLC3, ACY1, PROZ, GOT1, COL6A1, PSMB4, PEPD, BPGM, PSME1, IGLV3-21, CECR1, ALAD, PLA2G7, CPA1, PAFAH1B3, CTBS, BLVRB, FBP1, FUCA2, IGHG2, IGKC, PITHD1, B4GALT1, ALDH1A1

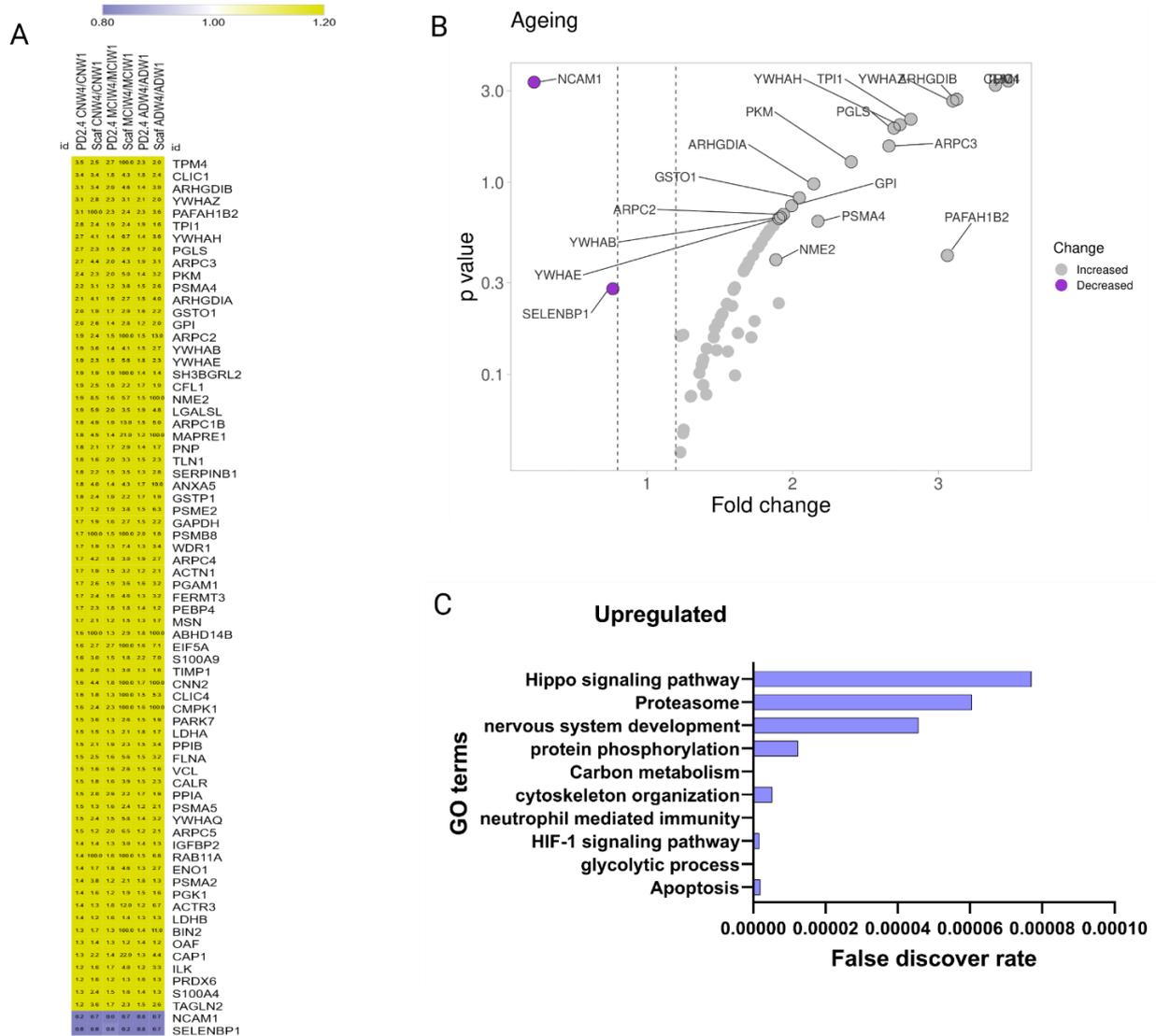
<p>MCI specific changes in <u>cross-sectional analysis</u> see Table 4.8 for details of fold change per group and p-value</p>	<p>Total protein number = 89 (53 Upregulated, 36 Downregulated)</p> <p>ITLN1, OGN, TRHDE, PRKACB, HSPA4, REG3A, ASGR2, NCAM1, IGF1, ORM1, GLOD4, APMAP, ANXA4, PPP1R7, SOD3, DUSP3, SULT1A1, LMAN2, PRDX4, CMPK1, IGFBP6, PON1, PRKAR1A, TF, C1QTNF3, C4BPB, PROC, IGFBP4, FAM3C;WNT16, A2M, IGFBP7, OAF, PEBP4, RARRES2, SNX3, CKM, APOL1, BST1, CR2, CDHR5, NIF3L1, PTPRF, CHI3L1, MASP2, C19orf10;MYDGF, ITIH3, ADAMTSL4, GANAB, CBR1;SETD4, COL5A1, MAN2A1, LTF, GPLD1, Sep-07, SELL, PPBP, GDI1, ARPC5, SPARC, CLIC1, FUCA2, CAT, NID1, LGALSL, SEPP1;SELENOP, FLT4, L1CAM, CPQ, PIP4K2A, PYGB, YWHAG, VWF, SVEP1, PSME1, MYH2, KRTAP4-4, CPA1, GSR, PEPD, BIN2, PA2G4, CECR1;ADA2, PPP2R4;PTPA, OTUB1, ICAM2, CALD1, LTBP1, AMY2A, KRTAP3-1</p>
<p>MCI specific changes common to both longitudinal and cross-sectional analysis see Table 4.11</p>	<p>Total protein number = 16</p> <p>PPBP, LTF, ASGR2, PEPD, CPA1, COL5A1, IGFBP6, HSPA4, PSME1, PPP1R7, ADAMTSL4, ITLN1, C19orf10, RARRES2, FUCA2, CECR1</p>
<p>AD specific changes in <u>longitudinal analysis</u> see Table 4.4 for details of fold change per group and p-value</p>	<p>Total protein number = 60 (39 Upregulated, 21 downregulated)</p> <p>TPM1, MST1L, CAPNS1, HPRT1, AMY1A, BID, PTPRK, S100A7, SERPINB9, HPR, PSMB8, SH3BGRL3, SDPR, GPX3, MAPRE2, OIT3, RAN, COL5A1, FAM3C, GLIPR2, PSMB2, UMOD, MGAT1, PA2G4, PAM, CYCS, VCP, QDPR, IGFBP6, CECR1;ADA2, ACTG1, ELTD1, ALDOB, LAMA2, APOD, IGFBP5, HSP90B1, IGLC3, TXNL1, ENDOD1, KRT35, LTF, CDH2, SERPINA1, SELL, KRT5, RNH1, KPRP, COL1A1, KRT6A, EGFR, KRT13, TNXB, ALDH1A1, ITLN1, MAN2A2, TF, KRT86, OLFML3, ADAMDEC1</p>
<p>AD specific changes in <u>cross-sectional analysis</u> see Table 4.7 for details of fold change per group and p-value</p>	<p>Total protein number = 70 (27 Upregulated, 43 downregulated)</p> <p>PITHD1, S100A7, PRDX1, PRDX4, CALD1, PSMB2, IGFBP1, TNC, COL6A1, MANBA, FAM3C, WNT16, RTN4RL2, F7, QDPR, GNPTG, PTPRK, CNTN3, PROZ, PAM, EXT2, NAPA, C1QTNF3, CHI3L1, ALDOB, CTSD, CFH, LAMA2, TF, NID1, PYGB, PPP2R4, PTPA, GANAB, LGALSL, BIN2, CLTC, PPBP, CLIC1, VCL, GP6, TNXB, PTPN6, ISOC1, GDI1, WARS, ECI1, DSP, CNN2, PGK1, TPI1, MAN2A2, VCP, SEPT7, CYCS, SPARC, PSMF1, PNP, ENG, CPB1, SND1, MAPRE2, ITGA2B, TYMP, CUTA, EGFR, RNASET2, PKP1, ACO1, FDPS, ARPC5, TUBB</p>
<p>AD specific changes common to both longitudinal and cross-sectional analysis see Table 4.10</p>	<p>Total protein number = 15</p> <p>EGFR, TF, ALDOB, QDPR, PAM, TNXB, LAMA2, S100A7, MAN2A2, PSMB2, VCP, CYCS, PTPRK, MAPRE2, FAM3C</p>
<p>Preclinical AD specific changes ADW1/CTRLW1 see Table 4.12 for details of fold change per group and p-value</p>	<p>Total protein number = 160 (59 Upregulated, 101 downregulated)</p> <p>PPP2R4;PTPA, KRTAP132, ITLN1, AMY2B, ORM1, SERPINA1, ALB, CRP, BLVRB, TTR, CHIT1, CAMP, CDH2, LECT2, KRT86, PSMB6, PRDX2, ALDH1A1, TF, C4BPB, CHI3L1, APOA1, AK1, CLEC3B, PSMB2, RNH1, PROCR, BPGM, PSMA4, SPP2, PCOLCE, PSMB1, C3, MFAP4, IGFALS, CFHR5, FCGBP, CD93, C1QB, C1orf68, MPO, PTGDS, F12, SELL, PLXND1, SOD2, LCAT, MBL2, LUM, TIMP2, KRT31, PARK7, PTPRS, LILRA3, IL6ST, C1QA, PSMA2, PEBP1, APOD, CORO1A, C19orf10;MYDGF, WDR1, ZG16,</p>

	MAPRE2, ALDOB, AHCY, PGK1, SERPINA11, PKM, PGD, LUZP6;MTPN, HSPA8, SEMG1, SEPT2, GANAB, FAH, PLEK, ACTN1, FABP4, EXT1, CRHBP, GSR, GAPDH, VASP, FUCA1, ISOC1, LDHA, YWHAH, CALR, VCL, SSC5D, ENO1, PCYOX1, UBA7, ZYX, ROBO4, F13A1, LTA4H, CSTB, P4HB, TUBB1, SH3BGRL2, CLIC4, TLN1, ALDOC, VCP, PF4, TWF2, GRB2, SVEP1, LAMC1, HYOU1, PDLIM1, TXN, PDIA6, ICOSLG;, LOC102723996, FERMT3, PFN1, HSP90AA1, SH3BGRL, UBA1, PDIA3, CNTN4, NUTF2, TGFB1, GMFG, CAPZA2, LPA, THBS1, AKR1A1, SPARC, FCGR2A, CAPZA1, CAP1, PPIB, CAPN1, APOC4;APOC4, APOC2, ELTD1;ADGRL4, COTL1, FCN1, ESD, SDPR;, CAVIN2, ARPC1B, COL5A1, PAM, LGALSL, ANGPTL3, MIF, YWHAQ, ARPC5, ARHGAP1, KRT36, ANXA5, RSU1, FABP1, ARPC2, DUSP3, PRDX5, SERPINB9, BID
Preclinical AD (ADW1/CTRLW1) common with incipient AD (ADW4/CTRLW4) see Table 4.13	Total protein number = 15 (4 Upregulated, 11 downregulated) ARPC5, PGK1, TF, ALDOB, SPARC, VCL, PAM, CHI3L1, PSMB2, VCP, GANAB, PPP2R4, PTPA, MAPRE2, LGALSL, ISOC1

4.3.2 DEPs identified in longitudinal analyses of ageing regardless of diagnosis

Comparing proteomic expression differences across the longitudinal cohorts provides insight into age-related changes, which are common across all three clinical groups, and appear to be largely independent of diagnosis. I observed that 71 proteins were dysregulated with ageing, the majority of which were upregulated (Figure 4.3iF, Figure 4.4A and Table 4.3). These 71 age-related DEPs were manually grouped into 12 protein functional categories based on gene ontology (GO) using the PD2.4 analysis outcomes (Figure 4.5A). The three functional groups with the highest number of age-related DEPs were cell signalling (35%), cytoskeleton and microtubules function (17%), and metabolism (15%) (Figure 4.5A). A variety of other categories represented $\leq 8\%$ of total DEPs each (Figure 4.5A). Of 71 DEPs in normal ageing, only two proteins were decreased, these being methanethiol oxidase (SELENBP1) and neuronal adhesion molecule 1 (NCAM1) Figure 4.4A. By contrast, proteins associated with inflammation (S100A9, S100A4, YWHA/14-3-3 family proteins), metabolic proteins (LDHA, LDHB, PKM, NME2), proteasome subunits (PSMA4, PSME2, PSMB8, PSMA6, PSMA5), and DNA binding and repair (ENO1, PARK7, CALR) were increased with ageing in all three clinical groups in the longitudinal analysis. However, they were not specific to disease (Figure 4.4 and Table 4.3). The complete list of proteins that are differentially expressed in ageing is shown in Table 4.3 and heatmap (Figure 4.4A), while a volcano plot shows the top 20 age-related DEPs with the greatest fold change (FC) Figure 4.4B.

Figure 4.4. **A:** Heatmap of 71 dysregulated proteins containing 69 upregulated and 2 downregulated in a similar direction in all normal ageing, MCI and AD showing the plasma proteome changes with age and not specific to the disease. **B:** Volcano plots highlight the 20 DEPs with the highest fold change in longitudinal ageing (I have highlighted only the top 20 proteins to avoid the overcrowding on volcano plots); the complete list of DEPs with age are presented in Table 4.3. **C:** This figure presents the Upregulated GO enrichment of pathways linked to ageing. However, only 2 DEPs were downregulated in ageing, no GO enrichment was identified in STRING software for downregulated proteins.



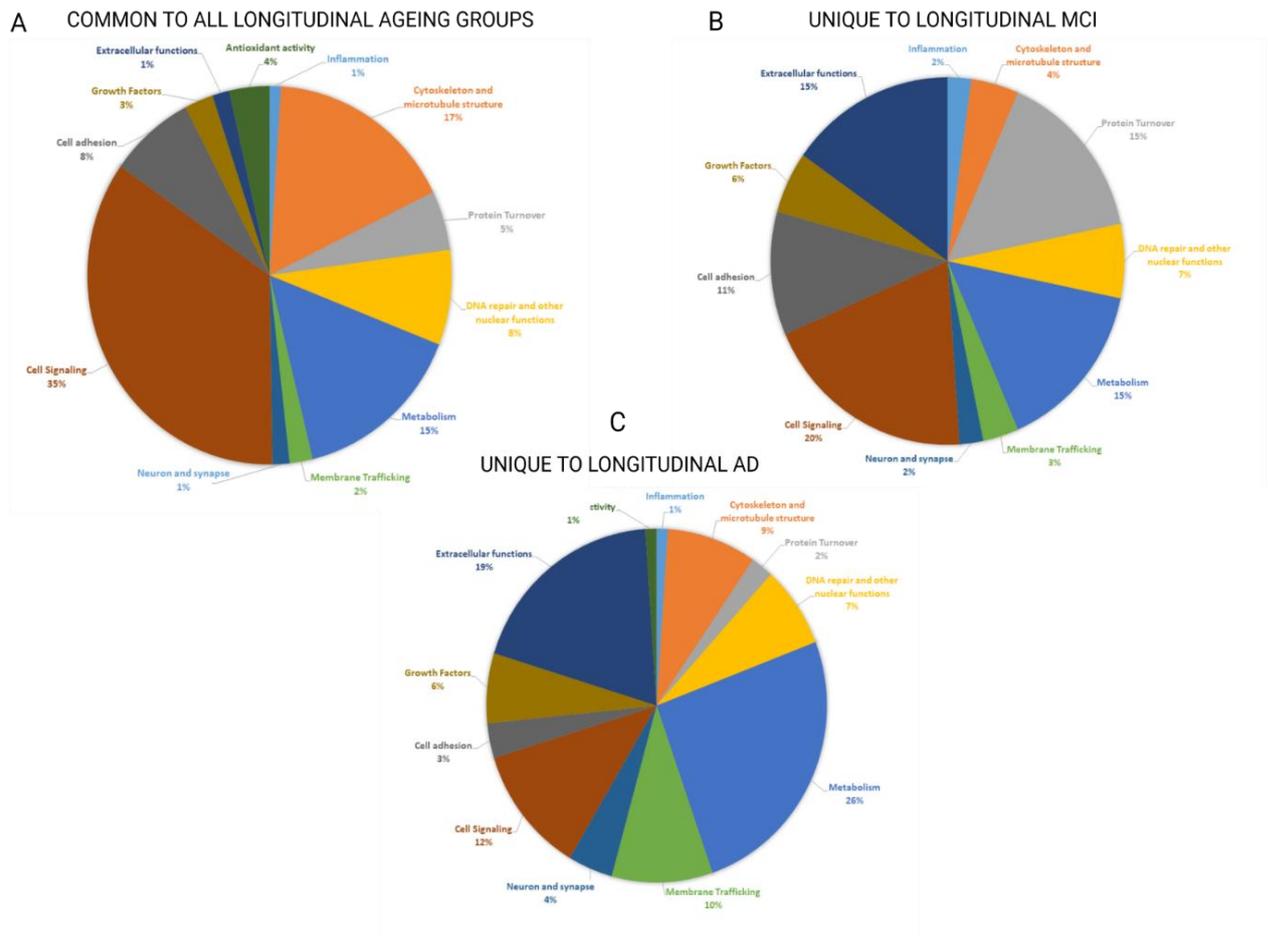


Figure 4.5. The final list of DEPs from longitudinal comparisons was sorted into lists based on the protein function classes, i.e., extracellular function, cell adhesion, growth factors, cell signalling, neuroinflammation, cytoskeleton, protein turnover, DNA binding repair, metabolism, membrane trafficking, neuron and synapse, antioxidant activity: **A.** longitudinal ageing, **B.** longitudinal MCI, **C.** longitudinal AD.

Several age-related DEPs with the highest fold change include the following: Tropomyosin alpha-4 (TPM4, FC= 3.4, p=0.00), chloride intracellular channel protein 1 (CLIC1, FC=3.391, p= 0.00), Rho GDP-dissociation inhibitor 2 (ARHGDI2, FC=3.126, p=0.00), 14-3-3 protein zeta/delta (YWHAZ, FC=3.09, p= 0.00), 6-phosphogluconolactonase (PGLS), Nucleoside diphosphate kinase B (NME2, FC=8.5, p=0.04), and NCAM1 (FC= 0.2, p=0.00) (see Table 4.3 for the full list). Gene ontology (GO) enrichment analysis of these 71 DEPs was performed to understand the molecular pathways affected in normal ageing (Table 4.3). The ageing related DEPs list was analyzed using the STRING bioinformatics tool, and enrichment in multiple GO-based categories was observed, including; 157 biological processes, 36 cellular components, 16 molecular functions, 48 KEGG and Reactome pathways.

Table 4.3: The final list of longitudinal 71 DEPs in normal ageing (changed in a similar direction in normal longitudinal ageing, MCI and AD were considered ageing-related changes, not specific to the disease).

Accession	Description	Gene Symbol	PD2.4_Abu ndance Ratio: (CTRLW4) / (CTRLW1)	Scaffold_Fol d change (CTRLW4) / (CTRLW1)	PD2.4_Abu ndance Ratio: (MCIW4) / (MCIW1)	Scaffold_Fol d change (MCIW4) / (MCIW1)	PD2.4_Abu ndance Ratio: (ADW4) / (ADW1)	Scaffold_Fol d change (ADW4) / (ADW1)
P31946	14-3-3 protein beta/alpha	YWHAB	1.917	3.6	1.37	4.1	1.502	2.7
P62258	14-3-3 protein epsilon	YWHAE	1.906	2.3	1.489	5.6	1.84	2.3
Q04917	14-3-3 protein	YWHAH	2.736	4.1	1.434	6.7	1.369	3.6
P27348	14-3-3 protein theta OS	YWHAQ	1.467	2.4	1.55	5.8	1.381	3.2
P63104	14-3-3 protein zeta/delta	YWHAZ	3.099	2.8	2.303	3.1	2.125	2
O95336	6- phosphogluconolact onase	PGLS	2.695	2.3	1.45	2.6	1.717	3
O15143	Actin-related protein 2/3 complex subunit 1B	ARPC1B	1.841	4.9	1.92	13	1.493	5
O15144	Actin-related protein 2/3 complex subunit 2	ARPC2	1.937	2.4	1.452	INF	1.536	13
O15145	Actin-related protein 2/3 complex subunit 3	ARPC3	2.661	4.4	2.028	4.3	1.949	3.1
P59998	Actin-related protein 2/3 complex subunit 4	ARPC4	1.686	4.2	1.833	3	1.945	2.7
O15511	Actin-related protein 2/3 complex subunit 5	ARPC5	1.459	1.2	1.96	6.5	1.227	2.1

P61158	Actin-related protein 3	ACTR3	1.377	1.3	1.833	12	1.201	6.7
Q01518	Adenylyl cyclase-associated protein 1	CAP1	1.253	2.2	1.425	22	1.285	4.4
P12814	Alpha-actinin-1	ACTN1	1.682	1.9	1.483	3.2	1.207	2.1
P06733	Alpha-enolase	ENO1	1.387	1.7	1.845	4.6	1.27	2.7
P08758	Annexin A5	ANXA5	1.764	4	1.361	4.3	1.747	10
Q9UBW5	Bridging integrator 2	BIN2	1.304	1.7	1.271	INF	1.389	11
Q99439	Calponin-2	CNN2	1.594	4.4	1.754	INF	1.714	INF
P27797	Calreticulin	CALR	1.491	1.8	1.618	3.9	1.485	2.3
O00299	Chloride intracellular channel protein 1	CLIC1	3.391	3.4	1.821	4.3	1.839	2.4
Q9Y696	Chloride intracellular channel protein 4	CLIC4	1.585	1.8	1.324	INF	1.486	5.3
P23528	Cofilin-1	CFL1	1.9	2.5	1.766	2.2	1.696	1.9
P63241	Eukaryotic translation initiation factor 5A-1	EIF5A	1.605	2.7	2.737	INF	1.622	7.1
Q86UX7	Fermitin family homolog 3	FERMT3	1.668	2.4	1.561	4.6	1.338	3.2
P21333	Filamin-A	FLNA	1.515	2.5	1.57	5.6	1.455	3.2
Q3ZCW2	Galectin-related protein	LGALSL	1.869	5.9	2.037	3.5	1.893	4.8
P06744	Glucose-6-phosphate isomerase	GPI	1.994	2.6	1.398	2.8	1.215	2
P78417	Glutathione S-transferase omega-1	GSTO1	2.046	1.9	1.666	2.9	1.556	2.2
P09211	Glutathione S-transferase P	GSTP1	1.764	2.4	1.893	2.2	1.702	1.9

P04406	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	1.729	1.9	1.625	2.7	1.536	2.2
P18065	Insulin-like growth factor-binding protein 2	IGFBP2	1.411	1.4	1.343	3	1.445	1.3
Q13418	Integrin-linked protein kinase	ILK	1.249	1.6	1.749	4	1.214	3.3
P30740	Leukocyte elastase inhibitor	SERPIN B1	1.767	2.2	1.478	3.5	1.309	2.8
P00338	L-lactate dehydrogenase A chain	LDHA	1.519	1.5	1.304	2.1	1.84	1.7
P07195	L-lactate dehydrogenase B chain	LDHB	1.361	1.2	1.57	1.4	1.291	1.3
P01033	Metalloproteinase inhibitor 1	TIMP1	1.598	2	1.34	3	1.336	1.6
Q13228	Methanethiol oxidase	SELENBP1	0.767	0.8	0.817	0.2	0.873	0.7
Q15691	Microtubule-associated protein RP/EB family member 1	MAPRE1	1.826	4.5	1.369	21	1.223	INF
P26038	Moesin	MSN	1.664	2.1	1.234	1.5	1.307	1.7
P13591	Neural cell adhesion molecule 1	NCAM1	0.227	0.7	0.01	0.7	0.784	0.7
P22392	Nucleoside diphosphate kinase B	NME2	1.885	8.5	1.606	5.7	1.464	INF
Q86UD1	Out at first protein homolog	OAF	1.301	1.4	1.283	1.2	1.413	1.2
P62937	Peptidyl-prolyl cis-trans isomerase A	PPIA	1.485	2	2.867	2.2	1.674	1.9

P23284	Peptidyl-prolyl cis-trans isomerase B	PPIB	1.518	2.1	1.911	2.3	1.536	3.4
P30041	Peroxiredoxin-6	PRDX6	1.23	1.6	1.218	1.3	1.599	1.3
Q96S96	Phosphatidylethanolamine-binding protein 4	PEBP4	1.666	2.3	1.818	1.8	1.37	1.2
P00558	Phosphoglycerate kinase 1	PGK1	1.384	1.6	1.247	1.9	1.504	1.6
P18669	Phosphoglycerate mutase 1	PGAM1	1.674	2.6	1.915	3.6	1.609	3.2
P68402	Platelet-activating factor acetylhydrolase IB subunit beta	PAFAH1B2	3.062	INF	2.275	2.4	2.308	3.6
Q9UL46	Proteasome activator complex subunit 2	PSME2	1.739	1.2	1.86	3.8	1.482	6.3
P25787	Proteasome subunit alpha type-2	PSMA2	1.387	3.8	1.212	2.1	1.781	1.3
P25789	Proteasome subunit alpha type-4	PSMA4	2.174	3.1	1.227	3.6	1.521	2.6
P28066	Proteasome subunit alpha type-5	PSMA5	1.479	1.3	1.602	2.4	1.248	2.1
P28062	Proteasome subunit beta type-8	PSMB8	1.717	INF	1.474	INF	1.954	1.8
Q96IU4	Protein ABHD14B	ABHD14B	1.625	INF	1.312	2.9	1.778	INF
P26447	Protein S100-A4	S100A4	1.25	2.4	1.477	1.6	1.365	1.3
P06702	Protein S100-A9	S100A9	1.604	3	1.505	1.8	2.24	7
Q99497	Protein/nucleic acid deglycase DJ-1	PARK7	1.55	3.6	1.306	2.6	1.47	1.9
P00491	Purine nucleoside phosphorylase	PNP	1.813	2.1	1.698	2.9	1.352	1.7

P14618	Pyruvate kinase PKM	PKM	2.402	2.3	2.049	5	1.442	3.2
P62491	Ras-related protein Rab-11A	RAB11A	1.408	INF	1.614	INF	1.464	6.8
P52565	Rho GDP- dissociation inhibitor 1	ARHGDI A	2.145	4.1	1.583	2.7	1.548	4
P52566	Rho GDP- dissociation inhibitor 2	ARHGDI B	3.126	3.4	2.864	4.6	1.363	3.9
Q9UJC5	SH3 domain- binding glutamic acid-rich-like protein 2	SH3BGR L2	1.905	1.9	1.877	INF	1.401	1.4
Q9Y490	Talin-1	TLN1	1.789	1.6	2.021	3.3	1.503	2.3
P37802	Transgelin-2	TAGLN2	1.23	3.6	1.66	2.3	1.545	2.6
P60174	Triosephosphate isomerase	TPI1	2.811	2.4	1.91	2.4	1.936	1.6
P67936	Tropomyosin alpha- 4 chain	TPM4	3.478	2.5	2.661	INF	2.268	2
P30085	UMP-CMP kinase	CMPK1	1.556	2.4	2.26	INF	1.645	INF
P18206	Vinculin	VCL	1.508	1.6	1.624	2.6	1.501	1.6
O75083	WD repeat- containing protein 1	WDR1	1.699	1.9	1.339	7.4	1.329	3.4

4.3.3 DEPs that change longitudinally with progression to MCI and AD from normal cognition:

A total of 60 DEPs were identified uniquely in the longitudinal AD group (progression from cognitively normal at W1 to AD at W4, 6 years later); 39 upregulated and 21 downregulated (Figure 4.3iF and Table 4.4). In longitudinal MCI, a total of 66 proteins were differentially expressed, with 41 upregulated and 25 downregulated (Figure 4.3iF, Table 4.5). Though the total number of dysregulated proteins is very similar in both conditions, only 19 appear in both MCI and AD (Figure 4.7A), and of these, only 9 have the same direction of fold-change, while the other 10 have opposite directions of fold-change (Figure 4.7B). Heatmaps based on differential protein abundance values from both search engines depicted overall reproducibility as well as individual protein expression profiles in AD and MCI in Figures 4.6A and 4.6C, respectively. Volcano plots highlight the top 20 DEPs with the greatest magnitude of longitudinal fold-change in AD and MCI (Figures 4.6B and 4.6D, respectively). Several DEPs unique to AD progression in W4, and that were significantly ($p \leq 0.05$) upregulated include: Tropomyosin alpha-1 chain (TPM1, fold change (FC)=19.4; $p=0.00$), Calpain small subunit 1 (CAPNS1, FC=2.6; $p=0.00$), Caveolae-associated protein 2 (SDPR, FC=18; $p=0.05$), Endoplasmin (HSP90B1, FC=1.5, $p=0.01$). Additionally, proteins that were significantly downregulated included: Alpha-mannosidase 2x (MAN2A2, FC=0.56; $p=0.02$), Olfactomedin-like protein 3 (OLFML3, FC=0.42; $p=0.03$), Keratin, type II cuticular Hb6 (KRT86, FC=0.51; $p=0.00$), and Serotransferrin (TF, FC=0.51; $p=0.00$). The complete list of DEPs unique to longitudinal AD group, is shown in Table 4.4. Functional categories with the greatest proportional change relative to either ageing or MCI, and with DEPs unique to longitudinal progression to AD, were associated with metabolism (26%), membrane trafficking (10%) and neuron & synapse (4%), all higher in AD than either control or MCI. In comparison, cell signalling (12%), cell adhesion (3%) and protein turnover (2%) are all lower in AD than either control or MCI (Figure 4.5B and 4.5C). The presence of proteins in plasma belonging to this AD progression specific groups implies functional disruptions which may have contributed to the progression of AD (Figure 4.5C). Two functional categories which were proportionately increased in both MCI and AD, relative to normal ageing were growth factors and extracellular functions (Figure 4.5). Their difference to normal ageing and common MCI and AD, suggests a possible association with cognitive impairment.

In AD (Table 4.4 DEP list), 39 upregulated proteins were associated with 36 biological processes, 19 cellular components, 12 molecular functions, and 7 KEGG & Reactome pathways. Approximately half of the proteins were linked to binding activity (protein binding, signalling receptor binding, and calcium ion binding), stress response, small molecule metabolic process, extracellular regions, and cytoplasm.

The plasma proteome profile of longitudinal progression to MCI contained several unique DEPs not shared by AD and normal ageing W4 vs W1 groups (Figure 4.6 C and D). In particular, the protein turnover group was proportionately higher in MCI (15%) than either the normal ageing group or AD, while cytoskeletal & microtubule structure was lower in MCI (4%) than either of the other groups (Figure 4.5B and C). MCI-specific DEPs with particularly high fold change with ageing, are shown in the Figure 4.6C and 4.6D heatmap and volcano plot and include upregulation of GTP-binding protein (RHEB, FC=46.691; p=0.00), pleckstrin (FC= 5.1; p=0.00), F-actin-capping protein subunit alpha-2 (CAPZA2, FC=25; p=0.00), insulin-like growth factor-binding protein 6 (IGFBP6, FC=1.8, p=0.01). Significantly downregulated proteins in MCI W4/W1 included flavin reductase NADPH (BLVRB, FC=0.4; p=0.02). The full list of DEPs in longitudinal MCI is shown in Tables 4.2 and 4.5.

In MCI, upregulated proteins from Table 4.5 were based on GO term enrichment analysis were significantly associated with 99 biological processes, 22 cellular components, 7 molecular functions, and 2 KEGG & Reactome pathways. Most DEPs fell into GO categories of; cellular process, cellular protein metabolism, regulation of protein phosphorylation, unfolded proteins, phosphate metabolic process, endomembrane system, signalling receptor binding and hemostasis. On the other hand, downregulated proteins from the MCI longitudinal analysis from Table 4.5 were significantly enriched in 33 biological processes, 10 molecular functions, and 11 KEGG & Reactome pathways.

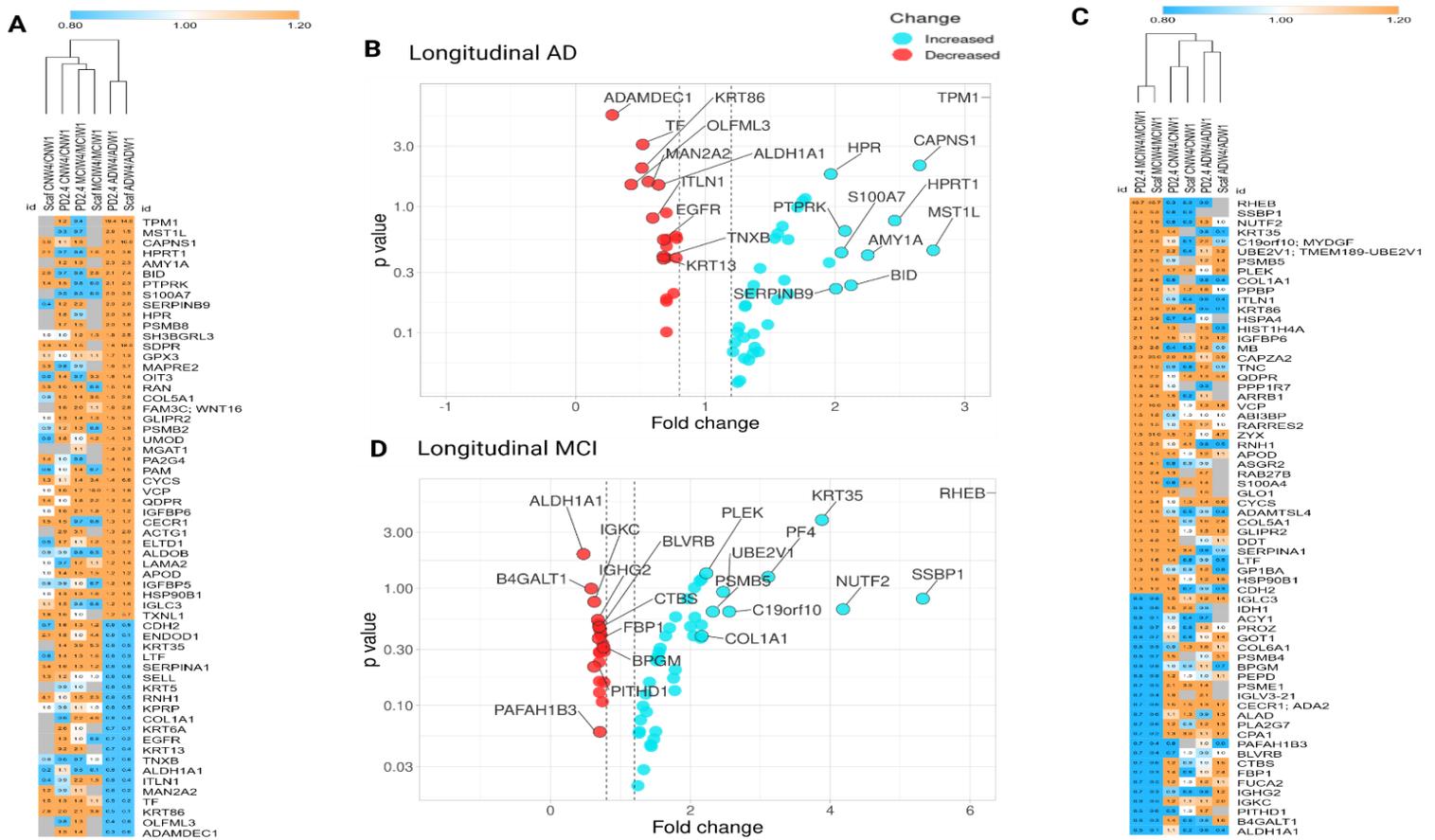


Figure 4.6. A and C. Heat map analysis of unique DEPs in longitudinal AD (ADW4/ADW1) and MCI (MCIW4/MCIW1) respectively. B and D: Volcano plots highlight the 20 DEPs with the highest fold change in the AD and MCI longitudinal analysis (I have highlighted only the top 20 proteins to avoid overcrowding on the volcano plots). The complete list of DEPs, including p values, is shown in Tables 4.4 and 4.5, respectively.

Table 4.4: The list of proteins uniquely differentially expressed in AD in longitudinal analysis (ADW4/ADW1).

Accession	Description	Gene Symbol	PD2.4_Abundance Ratio: (ADW4) / (ADW1)	Scaffold_Fold change (ADW4) / (ADW1)	PD2.4_Abundance Ratio Adj. P-Value: (ADW4) / (ADW1)	Scaffold_P value(ADW4) / (ADW1)
P09493	Tropomyosin alpha-1 chain	TPM1	19.405	14	5.36E-08	0.023
Q2TV78	Putative macrophage stimulating 1-like protein	MST1L	2.757	1.5	0.355113	0.67
P04632	Calpain small subunit 1	CAPNS1	2.652	16	0.007682	0.12
P00492	Hypoxanthine-guanine phosphoribosyltransferase	HPRT1	2.46	3.8	0.168748	0.2
P04745	Alpha-amylase 1	AMY1A	2.252	2.3	0.387844	0.45
P55957	BH3-interacting domain death agonist	BID	2.123	7.4	0.5792	0.17
Q15262	Receptor-type tyrosine-protein phosphatase kappa	PTPRK	2.076	2.3	0.229221	0.38
P31151	Protein S100-A7	S100A7	2.049	3.5	0.370187	0.5
P50453	Serpin B9	SERPINB9	2.006	2	0.598428	0.57
P00739	Haptoglobin-related protein	HPR	1.967	3.6	0.015629	0.21
P28062	Proteasome subunit beta type-8	PSMB8	1.954	1.8	0.437909	0.47
Q9H299	SH3 domain-binding glutamic acid-rich-like protein 3	SH3BGRL3	1.77	2.5	0.069793	0.049
O95810	Caveolae-associated protein 2	SDPR	1.75	18	0.079531	0.055
P22352	Glutathione peroxidase 3	GPX3	1.71	1.3	0.104229	0.13
Q15555	Microtubule-associated protein RP/EB family member 2	MAPRE2	1.639	3.7	0.627507	0.1
Q8WWZ8	Oncoprotein-induced transcript 3 protein	OIT3	1.639	1.4	0.284449	0.7
P62826	GTP-binding nuclear protein Ran	RAN	1.609	1.8	0.549037	0.38
P20908	Collagen alpha-1(V) chain	COL5A1	1.594	2.8	0.201157	0.05

Q92520	Protein FAM3C	FAM3C	1.555	2.8	0.656778	0.24
Q9H4G4	Golgi-associated plant pathogenesis-related protein 1	GLIPR2	1.536	1.3	0.276903	0.38
P49721	Proteasome subunit beta type-2	PSMB2	1.481	5.6	0.766017	0.081
P07911	Uromodulin	UMOD	1.422	1.3	0.475402	0.78
P26572	Alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	MGAT1	1.41	2.3	0.85012	0.61
Q9UQ80	Proliferation-associated protein 2G4	PA2G4	1.381	1.6	0.837954	0.68
P19021	Peptidyl-glycine alpha-amidating monooxygenase	PAM	1.372	1.5	0.579433	0.48
P99999	Cytochrome c	CYCS	1.369	6.6	0.853668	0.15
P55072	Transitional endoplasmic reticulum ATPase	VCP	1.347	1.8	0.623921	0.39
P09417	Dihydropteridine reductase	QDPR	1.334	5.4	0.870078	0.17
P24592	Insulin-like growth factor-binding protein 6	IGFBP6	1.31	1.2	0.684586	0.41
Q9NZK5	Adenosine deaminase 2	CECR1; ADA2	1.305	1.7	0.689796	0.19
P63261	Actin, cytoplasmic 2	ACTG1	1.297	2	0.809821	0.039
Q9HBW9	Adhesion G protein-coupled receptor L4	ELTD1	1.266	3.2	0.908269	0.4
P05062	Fructose-bisphosphate aldolase B	ALDOB	1.259	1.7	0.775233	0.12
P24043	Laminin subunit alpha-2	LAMA2	1.249	1.4	0.911405	0.7
P05090	Apolipoprotein D	APOD	1.247	1.2	0.79198	0.58
P24593	Insulin-like growth factor-binding protein 5	IGFBP5	1.228	1.6	0.824548	0.35
P14625	Endoplasmin	HSP90B1	1.213	1.5	0.85012	0.01
P0DOY3	Immunoglobulin lambda constant 3	IGLC3	1.201	1.4	0.865653	0.45

O43396	Thioredoxin-like protein 1	TXNL1	1.219	5.7	0.984986	0.16
O94919	Endonuclease domain-containing 1 protein	ENDOD1	0.882	0.5	0.791993	0.28
Q92764	Keratin, type I cuticular Ha5	KRT35	0.813	0.1	0.665053	0.22
P02788	Lactotransferrin	LTF	0.806	0.5	0.652507	0.15
P19022	Cadherin-2	CDH2	0.801	0.3	0.4066	0.34
P01009	Alpha-1-antitrypsin	SERPINA1	0.797	0.8	0.327383	0.91
P14151	L-selectin	SELL	0.78	0.6	0.280574	0.072
P13647	Keratin, type II cytoskeletal 5	KRT5	0.779	0.5	0.277712	0.011
P13489	Ribonuclease inhibitor	RNH1	0.776	0.5	0.4066	0.51
Q5T749	Keratinocyte proline-rich protein	KPRP	0.774	0.5	0.261166	0.061
P02452	Collagen alpha-1(I) chain	COL1A1	0.755	0.4	0.623921	0.58
P02538	Keratin, type II cytoskeletal 6A	KRT6A	0.699	0.7	0.128638	0.43
P00533	Epidermal growth factor receptor	EGFR	0.679	0.2	0.285206	0.19
P13646	Keratin, type I cytoskeletal 13	KRT13	0.677	0.4	0.412871	0.43
P22105	Tenascin-X	TNXB	0.676	0.8	0.396474	0.2
P00352	Retinal dehydrogenase 1	ALDH1A1	0.639	0.4	0.032948	0.44
Q8WWA0	Intelectin-1	ITLN1	0.594	0.4	0.154662	0.4
P49641	Alpha-mannosidase 2x	MAN2A2	0.56	0.2	0.026716	0.22
P02787	Serotransferrin	TF	0.517	0.2	0.000794	0.065
O43790	Keratin, type II cuticular Hb6	KRT86	0.511	0.07	0.009598	0.099
Q9NRN5	Olfactomedin-like protein 3	OLFML3	0.428	0.8	0.03225	0.84
O15204	ADAM DEC1	ADAMDEC1	0.282	0.6	4.95E-06	0.7

Table 4.5: The list of proteins uniquely differentially expressed in MCI in longitudinal analysis (MCIW4/MCIW1)

Accession	Description	Gene Symbol	PD2.4_Abundance Ratio: (MCIW4) / (MCIW1)	Scaffold_Fold change (MCIW4) / (MCIW1)	PD2.4_Abundance Ratio Adj. P-Value: (MCIW4) / (MCIW1)	Scaffold_P value(MCIW4) / (MCIW1)
Q15382	GTP-binding protein Rheb	RHEB	46.691	INF	3.56E-07	0.33
Q04837	Single-stranded DNA-binding protein, mitochondrial	SSBP1	5.324	INF	0.153119	1
P61970	Nuclear transport factor 2	NUTF2	4.183	1.9	0.217691	0.68
Q92764	Keratin, type I cuticular Ha5	KRT35	3.882	5.3	0.000149	0.43
Q969H8	Myeloid-derived growth factor	C19orf10	2.555	4.3	0.233602	0.22
Q13404	Ubiquitin-conjugating enzyme E2 variant 1	UBE2V1	2.469	7.3	0.117027	0.13
P28074	Proteasome subunit beta type-5	PSMB5	2.321	3.5	0.233602	0.36
P08567	Pleckstrin	PLEK	2.229	5.1	0.045428	0.0011
P02452	Collagen alpha-1(I) chain	COL1A1	2.163	4.6	0.406281	0.29
P02775	Platelet basic protein	PPBP	2.159	1.2	0.063703	0.76
Q8WWA0	Intellectin-1	ITLN1	2.158	1.5	0.326154	0.73
O43790	Keratin, type II cuticular Hb6	KRT86	2.148	3.8	0.422336	0.28
P34932	Heat shock 70 kDa protein 4	HSPA4	2.136	3.9	0.070025	0.24
P62805	Histone H4	HIST1H4A	2.065	1.4	0.2716	0.73
P24592	Insulin-like growth factor-binding protein 6	IGFBP6	2.058	1.8	0.097857	0.016
P02144	Myoglobin	MB	2.045	2.8	0.40466	0.1
P47755	F-actin-capping protein subunit alpha-2	CAPZA2	2.002	25	0.336535	0.0035
P24821	Tenascin	TNC	1.959	1.2	0.153829	0.56
P09417	Dihydropteridine reductase	QDPR	1.785	2.2	0.628449	0.52

Q15435	Protein phosphatase 1 regulatory subunit 7	PPP1R7	1.777	2.9	0.734765	0.54
P49407	Beta-arrestin-1	ARRB1	1.76	4.3	0.675052	0.3
P55072	Transitional endoplasmic reticulum ATPase	VCP	1.702	16	0.34793	0.036
Q7Z7G0	Target of Nesh-SH3	ABI3BP	1.646	1.8	0.406289	0.22
Q99969	Retinoic acid receptor responder protein 2	RARRES2	1.569	1.5	0.491953	0.23
Q15942	Zyxin	ZYX	1.549	31	0.521529	0.016
P13489	Ribonuclease inhibitor	RNH1	1.545	2.3	0.56624	0.43
P05090	Apolipoprotein D	APOD	1.524	1.5	0.55505	0.037
P07307	Asialoglycoprotein receptor 2	ASGR2	1.512	4.1	0.574602	0.19
O00194	Ras-related protein Rab-27B	RAB27B	1.506	2.4	0.871046	0.48
P26447	Protein S100-A4	S100A4	1.477	1.6	0.887897	0.65
Q04760	Lactoylglutathione lyase	GLO1	1.441	1.7	0.90245	0.65
P99999	Cytochrome c	CYCS	1.431	3.4	0.899122	0.12
Q6UY14	ADAMTS-like protein 4	ADAMTSL4	1.415	1.3	0.695326	0.49
P20908	Collagen alpha-1(V) chain	COL5A1	1.373	3.6	0.81655	0.056
Q9H4G4	Golgi-associated plant pathogenesis-related protein 1	GLIPR2	1.362	1.3	0.750275	0.39
P30046	D-dopachrome decarboxylase	DDT	1.333	4.8	0.936923	0.077
P01009	Alpha-1-antitrypsin	SERPINA1	1.323	1.2	0.79691	0.7
P02788	Lactotransferrin	LTF	1.288	1.6	0.841014	0.37
P07359	Platelet glycoprotein Ib alpha chain	GP1BA	1.272	1.3	0.871046	0.079
P14625	Endoplasmin	HSP90B1	1.267	1.6	0.875075	0.12
P19022	Cadherin-2	CDH2	1.25	1.2	0.953576	0.76
O75874	Isocitrate dehydrogenase [NADP] cytoplasmic	IDH1	0.874	0.6	0.743002	0.36
P0DOY3	Immunoglobulin lambda constant 3	IGLC3	0.843	0.8	0.690868	0.56

Q03154	Aminoacylase-1	ACY1	0.794	0.1	0.583165	0.14
P22891	Vitamin K-dependent protein Z	PROZ	0.771	0.7	0.521558	0.16
P17174	Aspartate aminotransferase, cytoplasmic	GOT1	0.77	0.7	0.517545	0.31
P12109	Collagen alpha-1(VI) chain	COL6A1	0.769	0.5	0.515771	0.15
P28070	Proteasome subunit beta type-4	PSMB4	0.765	0.7	0.69545	0.74
P12955	Xaa-Pro dipeptidase	PEPD	0.758	0.8	0.490273	0.6
P07738	Bisphosphoglycerate mutase	BPGM	0.758	0.8	0.487438	0.16
Q06323	Proteasome activator complex subunit 1	PSME1	0.748	0.5	0.461501	0.54
P80748	Immunoglobulin lambda variable 3-21	IGLV3-21	0.741	0.4	0.780722	0.45
Q9NZK5	Adenosine deaminase 2	CECR1	0.723	0.6	0.401063	0.31
P13716	Delta-aminolevulinic acid dehydratase	ALAD	0.721	0.6	0.397366	0.34
Q13093	Platelet-activating factor acetylhydrolase	PLA2G7	0.719	0.6	0.392089	0.16
P15085	Carboxypeptidase A1	CPA1	0.71	0.2	0.369326	0.091
Q15102	Platelet-activating factor acetylhydrolase IB subunit gamma	PAFAH1B3	0.707	0.4	0.872012	0.61
Q01459	Di-N-acetylchitobiase	CTBS	0.698	0.6	0.341392	0.12
P30043	Flavin reductase (NADPH)	BLVRB	0.698	0.4	0.339628	0.024
P09467	Fructose-1,6-bisphosphatase 1	FBP1	0.689	0.3	0.423045	0.35
Q9BTY2	Plasma alpha-L-fucosidase	FUCA2	0.689	0.5	0.317563	0.35
P01859	Immunoglobulin heavy constant gamma 2	IGHG2	0.678	0.3	0.289903	0.31
P01834	Immunoglobulin kappa constant	IGKC	0.626	0.5	0.171385	0.23
Q9GZP4	PITH domain-containing protein 1	PITHD1	0.624	0.6	0.610958	0.71
P15291	Beta-1,4-galactosyltransferase 1	B4GALT1	0.583	0.3	0.101291	0.27
P00352	Retinal dehydrogenase 1	ALDH1A1	0.473	0.07	0.010975	0.14

4.3.4 Common plasma proteome changes in longitudinal AD and MCI groups:

Only about 20% of total DEPs in AD and MCI longitudinal analysis groups were identified in both groups (Figure 4.7A, 19 DEPs). Of these 19 DEPs, 9 have the same direction of fold change; 8 are upregulated, and 1 are downregulated (Table 4.6). The other 10 DEPs changed in the opposite direction in MCI and AD, showing that at the molecular level, substantial differences are apparent between MCI and AD, in that not only are a majority of DEPs different between the two groups (Figure 4.7A and 4.7B) but that even a good proportion of the proteins identified in common in the two groups have different directions of change.

GO term enrichment analysis of the 19 DEPs shared by AD and MCI groups identified various functional groups, including metabolism, immune response, apoptosis, WNT signalling, and inflammation (Figure 4.7C). The two functional groups which have the greatest number of DEPs shared by both MCI and AD are metabolism, and immune response, suggesting that dysregulation of these two functions are shared between MCI and AD, while the majority of other DEPs are unique to each group (Figure 4.7A, 4.7B and 4.7C).

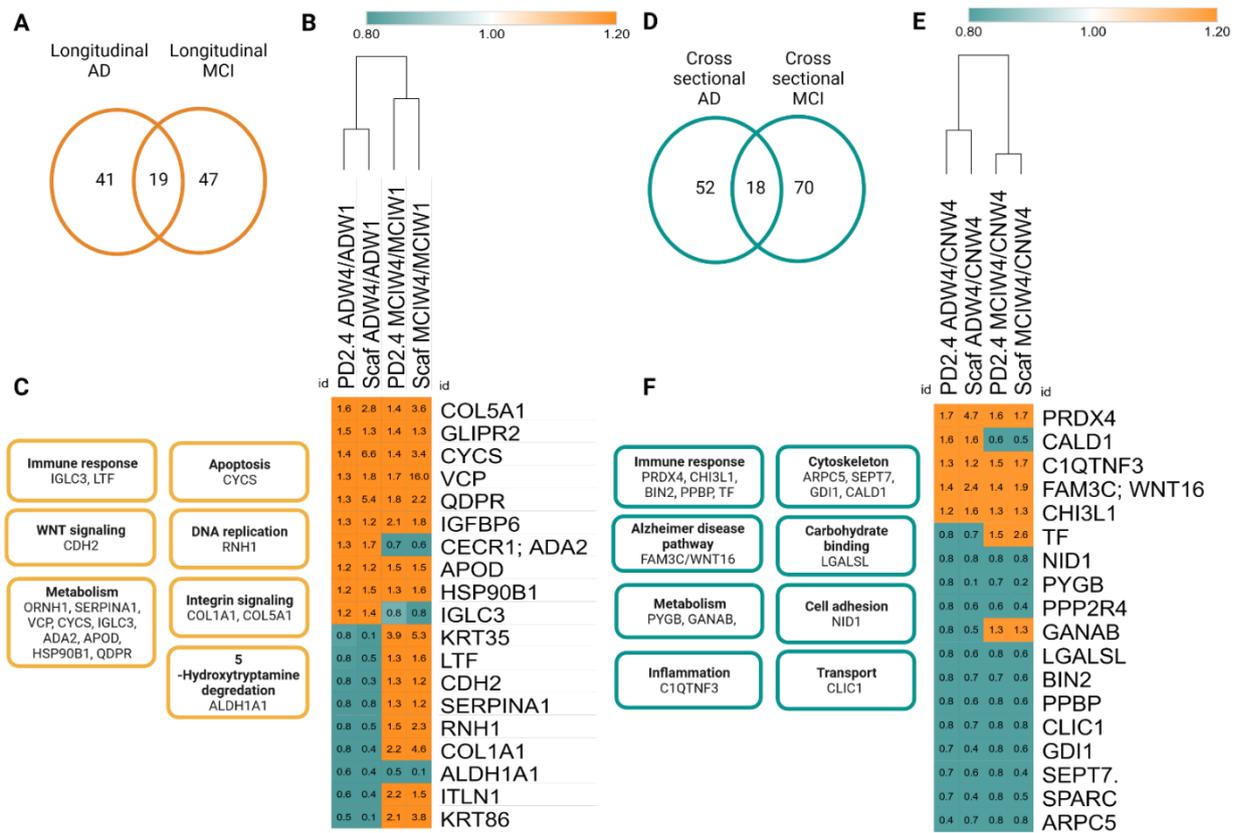


Figure 4.7: *A. Venn diagram showing 19 DEPs present in both longitudinal AD and MCI plasma proteome profiles (specific proteins are listed in Table 4.6). B. Heat map analysis of 19 DEPs common in longitudinal AD and MCI showing the pattern of common DEPs in both diseases. C. The 21 DEPs which were commonly dysregulated in MCI and AD were categorized into 10 GO enrichment terms including metabolism, immune response, apoptosis, WNT signalling, 5-Hydroxytryptamine degradation, a negative regulator of autophagy were associated with the list of common DEPs. D. Venn diagram showing 18 DEPs that were present in both cross-sectional AD and MCI plasma proteome profiles (Table 4.9). E. Heat map analysis of 18 DEPs common in cross sectional AD and MCI showing the pattern of common DEPs in both diseases. F. The cross-sectional common 18 DEPs were categorized into 9 GO enrichment includes an immune response, cytoskeleton, Alzheimer’s disease pathways, protein folding, metabolism, cell adhesion, inflammation, transport and carbohydrate-binding.*

Table 4.6: A total of 19 DEPs were common in both the longitudinal AD and MCI plasma proteome profiles

Accession	Description	Gene Symbol	PD2.4_Abundance Ratio: (ADW4) / (ADW1)	Scaffold_Fold change (ADW4) / (ADW1)	PD2.4_Abundance Ratio: (MCIW4) / (MCIW1)	Scaffold_Fold change (MCIW4) / (MCIW1)	PD2.4_Abundance Ratio Adj. P-Value: (ADW4) / (ADW1)	Scaffold_P value(ADW4) / (ADW1)	PD2.4_Abundance Ratio Adj. P-Value: (MCIW4) / (MCIW1)	Scaffold_P value(MCIW4) / (MCIW1)
P20908	Collagen alpha-1(V) chain	COL5A1	1.594	2.8	1.373	3.6	0.201	0.053	0.817	0.056
Q9H4G4	Golgi-associated plant pathogenesis-related protein 1	GLIPR2	1.536	1.3	1.362	1.3	0.277	0.38	0.750	0.39
P99999	Cytochrome c	CYCS	1.369	6.6	1.431	3.4	0.854	0.15	0.899	0.12
P55072	Transitional endoplasmic reticulum ATPase	VCP	1.347	1.8	1.702	16	0.624	0.39	0.348	0.036
P09417	Dihydropteridine reductase	QDPR	1.334	5.4	1.785	2.2	0.870	0.17	0.628	0.52
P24592	Insulin-like growth factor-binding protein 6	IGFBP6	1.31	1.2	2.058	1.8	0.685	0.41	0.098	0.016
Q9NZK5	Adenosine deaminase 2	CECR1; ADA2	1.305	1.7	0.723	0.6	0.690	0.19	0.401	0.31
P05090	Apolipoprotein D	APOD	1.247	1.2	1.524	1.5	0.792	0.58	0.555	0.037
P14625	Endoplasmic reticulum chaperone protein	HSP90B1	1.213	1.5	1.267	1.6	0.850	0.01	0.875	0.12
P0DOY3	Immunoglobulin lambda constant 3	IGLC3	1.201	1.4	0.843	0.8	0.866	0.45	0.691	0.56
Q92764	Keratin, type I cuticular Ha5	KRT35	0.813	0.1	3.882	5.3	0.665	0.22	0.000	0.43
P02788	Lactotransferrin	LTF	0.806	0.5	1.288	1.6	0.653	0.15	0.841	0.37
P19022	Cadherin-2	CDH2	0.801	0.3	1.25	1.2	0.407	0.34	0.954	0.76
P01009	Alpha-1-antitrypsin	SERPINA1	0.797	0.9	1.323	1.2	0.327	0.91	0.797	0.7

P13489	Ribonuclease inhibitor	RNH1	0.776	0.5	1.545	2.3	0.407	0.51	0.566	0.43
P02452	Collagen alpha-1(I) chain	COL1A1	0.755	0.4	2.163	4.6	0.624	0.58	0.406	0.29
P00352	Retinal dehydrogenase 1	ALDH1A1	0.639	0.4	0.473	0.07	0.033	0.44	0.011	0.14
Q8WWA0	Intelectin-1	ITLN1	0.594	0.4	2.158	1.5	0.155	0.4	0.326	0.73
O43790	Keratin, type II cuticular Hb6	KRT86	0.511	0.07	2.148	3.8	0.010	0.099	0.422	0.28

4.3.5 Cross sectional proteome changes in AD and MCI – potential clinical biomarkers:

Cross-sectional analyses compare the plasma proteome profiles of AD and MCI relative to their age-matched cognitively healthy controls. In the cross-sectional analysis of incipient AD group (ADW4/CTRLW4), 70 DEPs were identified, including 27 upregulated and 43 downregulated DEPs (Table 4.7). In MCI, 89 proteins were differentially expressed relative to age-matched normal controls (MCIW4/CTRLW4), with 53 upregulated and 36 downregulated DEPs (Table 4.8), indicating the number of dysregulated proteins identified in both disease conditions are similar. Heatmap analysis of the differentially abundant proteins in AD and MCI (Figure 4.8A and 4.8C respectively) show that there is some overlap of AD and MCI DEPs (Figure 4.7D). Volcano plots show the 20 DEPs in AD and MCI with the highest and lowest fold change in Figures 4.8B and 4.8D. Cross-sectional analyses of plasma proteome profiles of AD and MCI subjects relative to age-matched normal controls also identified a variety of potential disease-specific markers. DEPs identified in AD (ADW4/CTRLW4) that were not found in MCI (MCIW4/CTRLW4) (Figure 4.8A and 4.8B), including functions such as; antioxidants (PRDX4), proteasome (PSMB2, PSMF1), metabolism (MANBA, PYGB), cytoskeleton (TUBB, ARPC5). GO term enrichment analysis identified a diversity of significantly enriched categories in the DEPs upregulated in AD (9 biological processes, 14 cellular components, 7 molecular functions, 3 KEGG & Reactome pathways) and DEPs downregulated in AD (48 biological processes, 48 cellular components, 14 molecular functions, 2 KEGG & Reactome pathways). Approximately half of the proteins were associated with binding activity (protein binding, signalling receptor binding and calcium ion binding), response to stress, small molecule metabolic process, extracellular regions, and cytoplasm.

When compared to age-matched cognitively normal controls, the plasma proteome profile of MCI (MCIW4/CTRLW4) demonstrated a plethora of DEPs that were not observed in cross-sectional AD (ADW4/CTRLW4) group. These DEPs include functions such as growth factors (IGF1, MYDGF, OGN), metabolism (CBR1, GSR), signalling (YWHAG), immunity (ITLN1), and vascular function (VWF) (Figure 4.8C and 4.8D, Table 4.8). GO term enrichment analysis identified 76 biological processes, 27 cellular components, 11 molecular functions and 6 KEGG & Reactome pathway categories significantly enriched (Benjamini-Hochberg FDR <0.05) using DEPs upregulated in MCI. The main functional enrichments identified using the

DEPs upregulated in MCI included: metabolic process, vesicle-mediated transport, immune system process, homeostasis and the complement cascade. Enrichment analysis of DEPs downregulated in MCI identified 9 biological processes, 20 cellular components, 6 molecular functions, and 2 KEGG & Reactome pathway categories.

Relatively few of the same DEPs were identified in both the cross-sectional and longitudinal analyses of AD and MCI, being 15 (Table 4.10) and 16 (Table 4.11) DEPs, respectively.

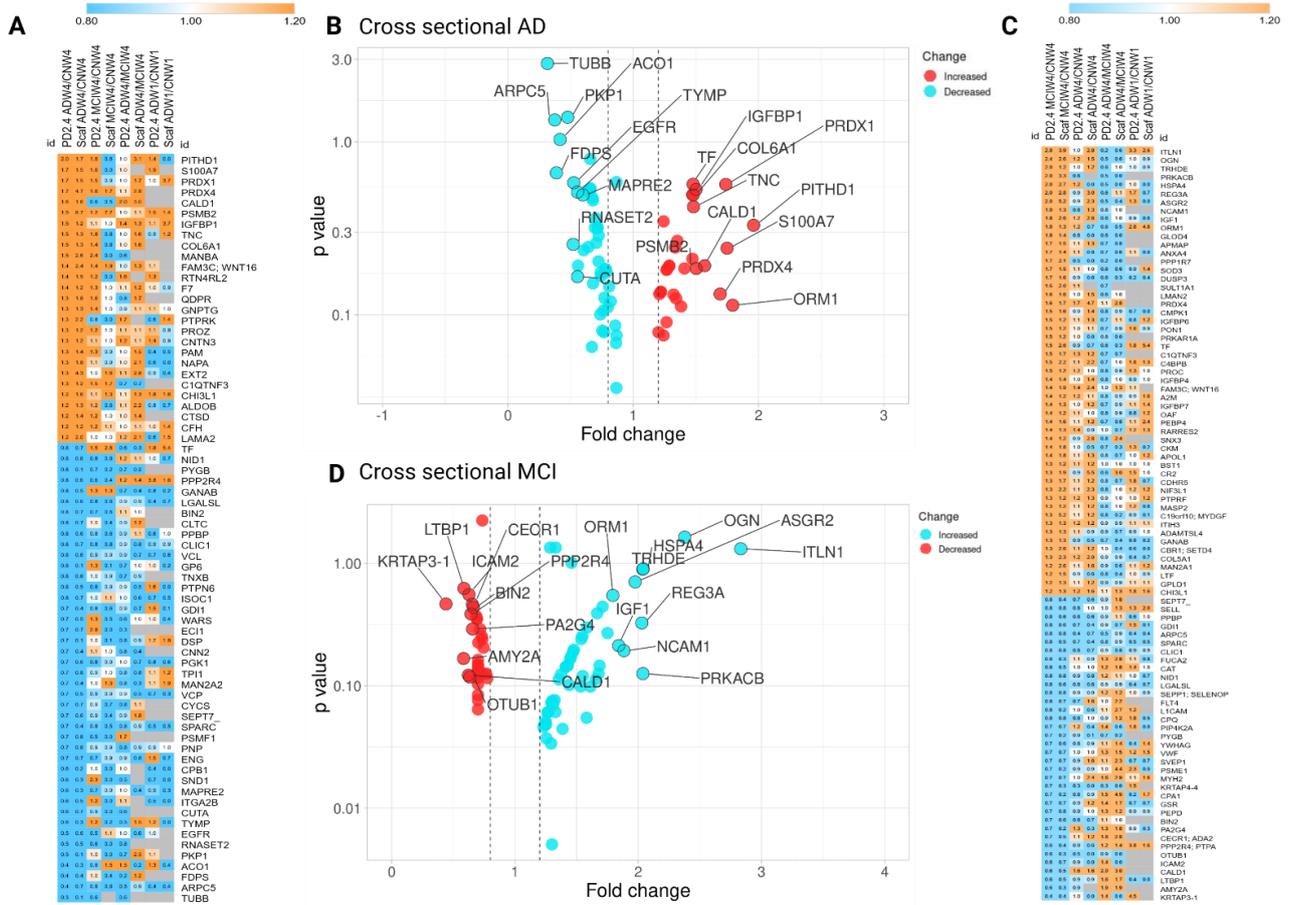


Figure 4.8. A and C Heat map analysis of DEPs in cross sectional comparisons of AD (ADW4/CTRLW4) and MCI (MCIW4/CTRLW4), respectively. C and D: Volcano plots highlight the 20 DEPs with the highest fold change in cross-sectional AD and MCI comparisons (I have highlighted only the top 20 proteins to avoid the overcrowding on volcano plots). A complete list of DEPs with p-values is shown in Table 4.7 and Table 4.8, respectively.

Table 4.7: The list of proteins uniquely differentially expressed in AD in cross-sectional comparison (ADW4/CTRLW4).

Accession	Description	Gene Symbol	PD2.4_Abundance Ratio: (ADW4) / (CTRLW4)	Scaffold fold change (ADW4) / (CTRLW4)	PD2.4_Abundance Ratio Adj. P-Value: (ADW4) / (CTRLW4)	Scaffold P-Value: (ADW4) / (CTRLW4)
Q9GZP4	PITH domain-containing protein 1	PITHD1	1.96	1.7	0.468578	0.66
P31151	Protein S100-A7	S100A7	1.749	1.5	0.572062	0.79
Q06830	Peroxiredoxin-1	PRDX1	1.739	1.5	0.270851	0.43
Q13162	Peroxiredoxin-4	PRDX4	1.694	4.7	0.738359	0.14
Q05682	Caldesmon	CALD1	1.57	1.6	0.642572	0.74
P49721	Proteasome subunit beta type-2	PSMB2	1.503	8.7	0.652877	0.083
P08833	Insulin-like growth factor-binding protein 1	IGFBP1	1.502	1.2	0.294789	0.65
P24821	Tenascin	TNC	1.481	1.3	0.379469	0.16
P12109	Collagen alpha-1(VI) chain	COL6A1	1.478	1.3	0.321531	0.59
O00462	Beta-mannosidase	MANBA	1.473	2.6	0.61575	0.47
Q92520	Protein FAM3C	FAM3C; WNT16	1.409	2.4	0.652877	0.32
Q86UN3	Reticulon-4 receptor-like 2	RTN4RL2	1.382	1.5	0.773128	0.72
P08709	Coagulation factor VII	F7	1.352	1.2	0.540064	0.26
P09417	Dihydropteridine reductase	QDPR	1.342	1.8	0.751176	0.51
Q9UJJ9	N-acetylglucosamine-1-phosphotransferase subunit gamma	GNPTG	1.335	1.3	0.568657	0.29
Q15262	Receptor-type tyrosine-protein phosphatase kappa	PTPRK	1.322	2.2	0.740136	0.45
Q9P232	Contactin-3	CNTN3	1.288	1.2	0.642572	0.65
P22891	Vitamin K-dependent protein Z	PROZ	1.288	1.2	0.642572	0.54
P19021	Peptidyl-glycine alpha-amidating monooxygenase	PAM	1.27	1.4	0.655308	0.59
Q93063	Exostosin-2	EXT2	1.266	4.3	0.658251	0.24
P54920	Alpha-soluble NSF attachment protein	NAPA	1.266	1.8	0.811847	0.48

Q9BXJ4	Complement C1q tumor necrosis factor-related protein 3	C1QTNF3	1.25	1.2	0.32	0.23994
P36222	Chitinase-3-like protein 1	CHI3L1	1.242	1.6	0.839637	0.18
P05062	Fructose-bisphosphate aldolase B	ALDOB	1.221	1.3	0.731685	0.45
P07339	Cathepsin D	CTSD	1.22	1.4	0.731685	0.47
P08603	Complement factor H	CFH	1.209	1.2	0.738399	0.14
P24043	Laminin subunit alpha-2	LAMA2	1.202	2	0.833246	0.5
P02787	Serotransferrin	TF	0.869	0.7	0.27	0.077121
P14543	Nidogen-1	NID1	0.867	0.8	0.839637	0.55
P11216	Glycogen phosphorylase, brain form	PYGB	0.865	0.06	0.916734	0.028
Q15257	Serine/threonine-protein phosphatase 2A activator	PPP2R4; PTPA	0.863	0.6	0.85397	0.4
Q14697	Neutral alpha-glucosidase AB	GANAB	0.857	0.5	0.819429	0.17
Q3ZCW2	Galectin-related protein	LGALSL	0.82	0.6	0.759523	0.15
Q9UBW5	Bridging integrator 2	BIN2	0.807	0.7	0.714901	0.47
Q00610	Clathrin heavy chain 1	CLTC	0.796	0.7	0.775769	0.59
P02775	Platelet basic protein	PPBP	0.787	0.6	0.667729	0.064
O00299	Chloride intracellular channel protein 1	CLIC1	0.781	0.7	0.655308	0.13
P18206	Vinculin	VCL	0.779	0.6	0.655308	0.024
Q9HCN6	Platelet glycoprotein VI	GP6	0.769	0.1	0.749127	0.094
P22105	Tenascin-X	TNXB	0.767	0.8	0.833246	0.12
P29350	Tyrosine-protein phosphatase non-receptor type 6	PTPN6	0.761	0.5	0.784524	0.4
Q96CN7	Isochorismatase domain-containing protein 1	ISOC1	0.753	0.7	0.83001	0.56
P31150	Rab GDP dissociation inhibitor alpha	GDI1	0.742	0.4	0.680185	0.16
P23381	Tryptophan--tRNA ligase, cytoplasmic	WARS	0.74	0.5	0.667564	0.24
P42126	Enoyl-CoA delta isomerase 1, mitochondrial	ECI1	0.737	0.7	0.79367	0.83
P15924	Desmoplakin	DSP	0.724	0.1	0.516532	0.29

Q99439	Calponin-2	CNN2	0.723	0.4	0.625472	0.059
P00558	Phosphoglycerate kinase 1	PGK1	0.716	0.6	0.48116	0.068
P60174	Triosephosphate isomerase	TPI1	0.715	0.8	0.478043	0.4
P49641	Alpha-mannosidase 2x	MAN2A2	0.706	0.4	0.549055	0.54
P55072	Transitional endoplasmic reticulum ATPase	VCP	0.704	0.5	0.435599	0.14
Q16181	Septin-7	SEPT7	0.679	0.6	0.347957	0.5
P99999	Cytochrome c	CYCS	0.679	0.7	0.705269	0.64
P09486	SPARC	SPARC	0.675	0.4	0.335137	0.0089
Q92530	Proteasome inhibitor PI31 subunit	PSMF1	0.67	0.8	0.860799	0.91
P00491	Purine nucleoside phosphorylase	PNP	0.659	0.8	0.292817	0.46
P17813	Endoglin	ENG	0.657	0.7	0.289361	0.77
P15086	Carboxypeptidase B	CPB1	0.642	0.2	0.566837	0.43
Q7KZF4	Staphylococcal nuclease domain-containing protein 1	SND1	0.605	0.3	0.580919	0.45
Q15555	Microtubule-associated protein RP/EB family member 2	MAPRE2	0.598	0.3	0.321531	0.0086
P08514	Integrin alpha-IIb	ITGA2B	0.559	0.5	0.642572	0.57
P19971	Thymidine phosphorylase	TYMP	0.555	0.3	0.307284	0.088
O60888	Protein CutA	CUTA	0.555	0.7	0.682359	0.83
P00533	Epidermal growth factor receptor	EGFR	0.525	0.6	0.262896	0.73
O00584	Ribonuclease T2	RNASET2	0.521	0.5	0.555491	0.53
Q13835	Plakophilin-1	PKP1	0.477	0.09	0.04102	0.33
P21399	Cytoplasmic aconitate hydratase	ACO1	0.416	0.3	0.092928	0.48
P14324	Farnesyl pyrophosphate synthase	FDPS	0.387	0.4	0.217861	0.34
O15511	Actin-related protein 2/3 complex subunit 5	ARPC5	0.374	0.7	0.045864	0.44
P07437	Tubulin beta chain	TUBB	0.315	0.1	0.001446	0.0087

Table 4.8: The list of proteins uniquely differentially expressed in MCI cross-sectional comparison (MCIW4/CTRLW4).

Accession	Description	Gene Symbol	PD2.4_Abundance Ratio: (MCIW4) / (CTRLW4)	Scaffold fold change: (MCIW4) / (CTRLW4)	PD2.4_Abundance Ratio Adj. P-Value: (MCIW4) / (CTRLW4)	Scaffold P-Value: (MCIW4) / (CTRLW4)
Q8WWA0	Intelectin-1	ITLN1	2.832	3.9	0.0481223	0.36
P20774	Mimecan	OGN	2.377	2.6	0.023	1.6382722
Q9UKU6	Thyrotropin-releasing hormone-degrading ectoenzyme	TRHDE	2.04	1.2	0.125004	0.903076
P22694	cAMP-dependent protein kinase catalytic subunit beta	PRKACB	2.038	3.3	0.7484872	0.1258156
P34932	Heat shock 70 kDa protein 4	HSPA4	2.037	2.7	0.1266229	0.8974876
Q06141	Regenerating islet-derived protein 3-alpha	REG3A	2.029	2.8	0.4717034	0.326331
P07307	Asialoglycoprotein receptor 2	ASGR2	1.975	5.2	0.1967669	0.706048
P13591	Neural cell adhesion molecule 1	NCAM1	1.884	1.3	0.6401493	0.1937187
P05019	Insulin-like growth factor I	IGF1	1.841	2.6	0.6123379	0.2130088
P02763	Alpha-1-acid glycoprotein 1	ORM1	1.794	1.3	0.2831411	0.547997
Q9HC38	Glyoxalase domain-containing protein 4	GLOD4	1.751	1.4	0.5371733	0.2698856
Q9HDC9	Adipocyte plasma membrane-associated protein	APMAP	1.711	1.5	0.3595098	0.4442893
P09525	Annexin A4	ANXA4	1.686	1.4	0.7141254	0.1462255
Q15435	Protein phosphatase 1 regulatory subunit 7	PPP1R7	1.678	2.1	0.7484872	0.1258156
P08294	Extracellular superoxide dismutase [Cu-Zn]	SOD3	1.667	1.5	0.4060378	0.3914335
P51452	Dual specificity protein phosphatase 3	DUSP3	1.657	1.6	0.7702635	0.1133607
P50225	Sulfotransferase 1A1	SULT1A1	1.612	2	0.7973053	0.0983754
Q12907	Vesicular integral-membrane protein VIP36	LMAN2	1.586	1.6	0.501516	0.2997152

Q13162	Peroxiredoxin-4	PRDX4	1.581	1.7	0.8814668	0.61
P30085	UMP-CMP kinase	CMPK1	1.55	1.6	0.7577868	0.120453
P24592	Insulin-like growth factor-binding protein 6	IGFBP6	1.548	1.2	0.5550578	0.2556618
P27169	Serum paraoxonase/arylesterase 1	PON1	1.537	1.2	0.5742667	0.2408863
P10644	cAMP-dependent protein kinase type I-alpha regulatory subunit	PRKAR1A	1.532	1.2	0.7969063	0.0985927
P02787	Serotransferrin	TF	1.478	2.6	0.6348591	0.27
Q9BXJ4	Complement C1q tumor necrosis factor-related protein 3	C1QTNF3	1.471	1.7	0.6459608	0.1897938
P20851	C4b-binding protein beta chain	C4BPB	1.458	2.2	0.097	1.0132283
P04070	Vitamin K-dependent protein C	PROC	1.454	1.2	0.6718201	0.172747
P22692	Insulin-like growth factor-binding protein 4	IGFBP4	1.448	1.4	0.6816923	0.1664116
Q92520	Protein FAM3C	FAM3C; WNT16	1.44	1.9	0.7913147	0.43
P01023	Alpha-2-macroglobulin	A2M	1.428	1.2	0.7119065	0.147577
Q16270	Insulin-like growth factor-binding protein 7	IGFBP7	1.426	1.3	0.7141254	0.1462255
Q86UD1	Out at first protein homolog OS	OAF	1.424	1.2	0.7141254	0.1462255
Q96S96	Phosphatidylethanolamine-binding protein 4	PEBP4	1.411	1.6	0.7201387	0.1425838
Q99969	Retinoic acid receptor responder protein 2	RARRES2	1.411	1.3	0.7201387	0.41
O60493	Sorting nexin-3	SNX3	1.386	1.2	0.9024726	0.044566
P06732	Creatine kinase M-type	CKM	1.375	1.8	0.75139	0.1241346
O14791	Apolipoprotein L1	APOL1	1.355	1.8	0.7708556	0.113027
Q10588	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 2	BST1	1.331	1.2	0.045	1.3467875
P20023	Complement receptor type 2	CR2	1.328	1.9	0.8687855	0.0610874

Q9HBB8	Cadherin-related family member 5	CDHR5	1.322	1.7	0.8383456	0.0765769
Q9GZT8	NIF3-like protein 1	NIF3L1	1.309	2.2	0.8516682	0.0697295
P10586	Receptor-type tyrosine-protein phosphatase F	PTPRF	1.302	1.2	0.8410053	0.0752013
P36222	Chitinase-3-like protein 1	CHI3L1	1.2	1.3	0.988409	0.0050633
O00187	Mannan-binding lectin serine protease 2	MASP2	1.295	1.2	0.845749	0.0727585
Q969H8	Myeloid-derived growth factor	C19orf10; MYDGF	1.293	5.2	0.9252289	0.0337508
Q06033	Inter-alpha-trypsin inhibitor heavy chain H3	ITIH3	1.285	1.3	0.045	1.3467875
Q6UY14	ADAMTS-like protein 4	ADAMTSL4	1.268	1.3	0.8687855	0.0610874
Q14697	Neutral alpha-glucosidase AB	GANAB	1.264	1.3	0.8697649	0.0605981
P16152	Carbonyl reductase [NADPH] 1	CBR1; SETD4	1.255	2.6	0.9178352	0.0372353
P20908	Collagen alpha-1(V) chain	COL5A1	1.253	2.3	0.8947421	0.0483021
Q16706	Alpha-mannosidase 2	MAN2A1	1.248	2.6	0.8877724	0.0516984
P02788	Lactotransferrin	LTF	1.242	1.5	0.8944913	0.0484239
P80108	Phosphatidylinositol-glycan-specific phospholipase D	GPLD1	1.231	1.3	0.8990445	0.0462188
Q16181	Septin-7	Sep-07	0.886	0.4	0.8627009	0.12
P14151	L-selectin	SELL	0.866	0.8	0.8383456	0.26
P02775	Platelet basic protein	PPBP	0.84	0.6	0.787821	0.065
P31150	Rab GDP dissociation inhibitor alpha	GDI1	0.829	0.6	0.8237697	0.3
O15511	Actin-related protein 2/3 complex subunit 5	ARPC5	0.823	0.8	0.7543202	0.6
P09486	SPARC	SPARC	0.822	0.5	0.752345	0.0074
O00299	Chloride intracellular channel protein 1	CLIC1	0.809	0.8	0.7337874	0.46

Q9BTY2	Plasma alpha-L-fucosidase	FUCA2	0.804	0.3	0.7216359	0.034
P04040	Catalase	CAT	0.796	0.5	0.7170453	0.1444534
P14543	Nidogen-1	NID1	0.794	0.8	0.7141254	0.1462255
Q3ZCW2	Galectin-related protein	LGALSL	0.793	0.6	0.7344685	0.1340268
P49908	Selenoprotein P	SEPP1; SELENOP	0.79	0.8	0.7097972	0.1488657
P35916	Vascular endothelial growth factor receptor 3	FLT4	0.783	0.7	0.6873885	0.1627977
P32004	Neural cell adhesion molecule L1 OS	L1CAM	0.776	0.2	0.7662757	0.1156149
Q9Y646	Carboxypeptidase Q	CPQ	0.766	0.8	0.7484872	0.1258156
P48426	Phosphatidylinositol 5-phosphate 4-kinase type-2 alpha	PIP4K2A	0.745	0.7	0.6236173	0.2050819
P11216	Glycogen phosphorylase, brain form	PYGB	0.743	0.2	0.7484872	0.1258156
P61981	14-3-3 protein gamma	YWHAG	0.736	0.6	0.5746125	0.240625
P04275	von Willebrand factor	VWF	0.735	0.7	0.0057	2.2441251
Q4LDE5	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	SVEP1	0.73	0.7	0.5562921	0.2546971
Q06323	Proteasome activator complex subunit 1	PSME1	0.717	0.2	0.5121104	0.2906364
Q9UKX2	Myosin-2	MYH2	0.703	0.7	0.5949192	0.225542
Q9BYR3	Keratin-associated protein 4-4	KRTAP4-4	0.694	0.3	0.7657821	0.48
P15085	Carboxypeptidase A1	CPA1	0.692	0.2	0.4451852	0.3514593
P00390	Glutathione reductase, mitochondrial	GSR	0.691	0.7	0.4416225	0.3549488
P12955	Xaa-Pro dipeptidase	PEPD	0.685	0.8	0.4253025	0.3713021
Q9UBW5	Bridging integrator 2	BIN2	0.663	0.6	0.3651087	0.4375779
Q9UQ80	Proliferation-associated protein 2G4	PA2G4	0.656	0.2	0.5114573	0.16

Q9NZK5	Adenosine deaminase 2	CECR1; ADA2	0.654	0.5	0.3439083	0.4635574
Q15257	Serine/threonine-protein phosphatase 2A activator	PPP2R4; PTPA	0.643	0.4	0.4104502	0.3867396
Q96FW1	Ubiquitin thioesterase OTUB1	OTUB1	0.634	0.3	0.7624006	0.3
P13598	Intercellular adhesion molecule 2	ICAM2	0.627	0.7	0.2777685	0.556317
Q05682	Caldesmon	CALD1	0.622	0.5	0.7550198	0.6
Q14766	Latent-transforming growth factor beta-binding protein 1	LTBP1	0.586	0.5	0.2368549	0.6255176
P04746	Pancreatic alpha-amylase	AMY2A	0.583	0.3	0.6804248	0.1672199
Q9BYR8	Keratin-associated protein 3-1	KRTAP3-1	0.44	0.4	0.3415649	0.4665268

4.3.6 Common proteome changes in cross sectional analysis of AD and MCI:

There were 18 DEPs common to both AD and MCI in the cross-sectional analyses (Figure 4.7D, E, and F and Table 4.9). These MCI and AD shared DEPs were associated with functions such as immune system (PRDX4, CHI3L1, BIN2, PPBP, TF), cytoskeleton (ARPC5, SEPT7, GDI1, CALD1) and metabolism (PYGB, GANAB). Only 4 of these upregulated DEPs had a similar direction of fold-change in both AD and MCI (PRDX4, CHI3L1, FAM3C, C1QTNF3), while opposite directions of fold change were observed for TF, GANAB, and CALD1, Figure 4.7E. The majority of DEPs common to both MCI and AD in the cross-sectional analyses were downregulated (11/18 proteins), Figure 4.7E and, Table 4.9.

Table 4.9. The list of 18 DEPs common to both the cross sectional AD and MCI analyses.

Accession	Description	Gene Symbol	PD2.4_Abundance Ratio: (ADW4) / (CTRLW4)	Scaffold fold change (ADW4) / (CTRLW4)	PD2.4_Abundance Ratio: (MCIW4) / (CTRLW4)	Scaffold fold change: (MCIW4) / (CTRLW4)	PD2.4_Abundance Ratio Adj. P-Value: (ADW4) / (CTRLW4)	Scaffold P-Value: (ADW4) / (CTRLW4)	PD2.4_Abundance Ratio Adj. P-Value: (MCIW4) / (CTRLW4)	Scaffold P-Value: (MCIW4) / (CTRLW4)
Q13162	Peroxiredoxin-4	PRDX4	1.694	4.7	1.581	1.7	0.738	0.14	0.881	0.61
Q05682	Caldesmon	CALD1	1.57	1.6	0.622	0.5	0.643	0.74	0.755	0.6
Q9BXJ4	Complement C1q tumor necrosis factor-related protein 3	C1QTNF3	1.25	1.2	1.471	1.7	0.32	0.24	0.646	0.19
Q92520	Protein FAM3C	FAM3C; WNT16	1.409	2.4	1.44	1.9	0.653	0.32	0.791	0.43
P36222	Chitinase-3-like protein 1	CHI3L1	1.242	1.6	1.3	1.3	0.84	0.18	0.988	0.005
P02787	Serotransferrin	TF	0.869	0.7	1.478	2.6	0.27	0.077	0.635	0.27
P14543	Nidogen-1	NID1	0.867	0.8	0.794	0.8	0.84	0.55	0.714	0.146
P11216	Glycogen phosphorylase, brain form	PYGB	0.865	0.06	0.743	0.2	0.917	0.028	0.748	0.126
Q15257	Serine/threonine-protein phosphatase 2A activator	PPP2R4	0.863	0.6	0.643	0.4	0.854	0.4	0.41	0.387
Q14697	Neutral alpha-	GANAB	0.857	0.5	1.264	1.3	0.819	0.17	0.87	0.061

	glucosidase AB									
Q3ZCW2	Galectin- related protein	LGALS1	0.82	0.6	0.793	0.6	0.76	0.15	0.734	0.134
Q9UBW5	Bridging integrator 2	BIN2	0.807	0.7	0.663	0.6	0.715	0.47	0.365	0.438
P02775	Platelet basic protein	PPBP	0.787	0.6	0.84	0.6	0.668	0.064	0.788	0.065
O00299	Chloride intracellular channel protein 1	CLIC1	0.781	0.7	0.809	0.8	0.655	0.13	0.734	0.46
P31150	Rab GDP dissociation inhibitor alpha	GDI1	0.742	0.4	0.829	0.6	0.68	0.16	0.824	0.3
Q16181	Septin-7	SEPT7	0.679	0.6	0.886	0.4	0.348	0.5	0.863	0.12
P09486	SPARC	SPARC	0.675	0.4	0.822	0.5	0.335	0.009	0.752	0.007
O15511	Actin-related protein ^{2/3} complex subunit 5	ARPC5	0.374	0.7	0.823	0.8	0.046	0.44	0.754	0.6

Table 4.10. The list of DEPs common to both longitudinal and cross sectional AD comparisons.

Accession	Description	Gene Symbol	PD2.4_Abu ndance Ratio: (ADW4) / (ADW1)	Scaffold Fold change (ADW4) / (ADW1)	PD2.4_Abund ance Ratio: (ADW4) / (CTRLW4)	Scaffol d fold change (ADW 4) / (CTRL W4)	PD2.4_Abund ance Ratio Adj. P-Value: (ADW4) / (ADW1)	Scaffold_ P value(AD W4) / (ADW1)	PD2.4_Abund ance Ratio Adj. P-Value: (ADW4) / (CTRLW4)	Scaffol d P- Value: (ADW 4) / (CTRL W4)
P00533	Epidermal growth factor receptor	EGFR	0.679	0.2	0.525	0.6	0.285	0.19	0.263	0.73
P02787	Serotransferrin	TF	0.517	0.2	1.478	2.6	0.001	0.065	0.270	0.08
P05062	Fructose-bisphosphate aldolase B	ALDOB	1.259	1.7	1.221	1.3	0.775	0.12	0.732	0.45
P09417	Dihydropteridine reductase	QDPR	1.334	5.4	1.342	1.8	0.870	0.17	0.751	0.51
P19021	Peptidyl-glycine alpha-amidating monooxygenase	PAM	1.372	1.5	1.27	1.4	0.579	0.48	0.655	0.59
P22105	Tenascin-X	TNXB	0.676	0.8	0.767	0.8	0.396	0.2	0.833	0.12
P24043	Laminin subunit alpha-2	LAMA2	1.249	1.4	1.202	2	0.911	0.7	0.833	0.50
P31151	Protein S100-A7	S100A7	2.049	3.5	1.749	1.5	0.370	0.5	0.572	0.79
P49641	Alpha-mannosidase 2x	MAN2A2	0.56	0.2	0.706	0.4	0.027	0.22	0.549	0.54

P49721	Proteasome subunit beta type-2	PSMB2	1.481	5.6	1.503	8.7	0.766	0.081	0.653	0.08
P55072	Transitional endoplasmic reticulum ATPase	VCP	1.347	1.8	0.704	0.5	0.624	0.39	0.436	0.14
P99999	Cytochrome c	CYCS	1.369	6.6	0.679	0.7	0.854	0.15	0.705	0.64
Q15262	Receptor-type tyrosine-protein phosphatase kappa	PTPRK	2.076	2.3	1.322	2.2	0.229	0.38	0.740	0.45
Q15555	Microtubule-associated protein RP/EB family member 2	MAPRE2	1.639	3.7	0.598	0.3	0.628	0.1	0.322	0.01
Q92520	Protein FAM3C	FAM3C	1.555	2.8	1.409	2.4	0.657	0.24	0.653	0.32

Table 4.11. The list of DEPs was common to both longitudinal and cross-sectional MCI comparisons.

Accession	Description	Gene Symbol	PD2.4_Abundance Ratio: (MCIW4) / (MCIW1)	Scaffold_Fold change (MCIW4) / (MCIW1)	PD2.4_Abundance Ratio: (MCIW4) / (CTRLW4)	Scaffold fold change: (MCIW4) / (CTRLW4)	PD2.4_Abundance Ratio Adj. P-Value: (MCIW4) / (MCIW1)	Scaffold_P value(MCIW4) / (MCIW1)	PD2.4_Abundance Ratio Adj. P-Value: (MCIW4) / (CTRLW4)	scaffold P-Value: (MCIW4) / (CTRLW4)
P02775	Platelet basic protein	PPBP	2.159	1.2	0.84	0.6	0.064	0.760	0.788	0.065
P02788	Lactotransferrin	LTF	1.288	1.6	1.242	1.5	0.841	0.370	0.894	0.048
P07307	Asialoglycoprotein receptor 2	ASGR2	1.512	4.1	1.975	5.2	0.575	0.190	0.197	0.706
P12955	Xaa-Pro dipeptidase	PEPD	0.758	0.8	0.685	0.8	0.490	0.600	0.425	0.371
P15085	Carboxypeptidase A1	CPA1	0.71	0.2	0.692	0.2	0.369	0.091	0.445	0.351
P20908	Collagen alpha-1(V) chain	COL5A1	1.373	3.6	1.253	2.3	0.817	0.056	0.895	0.048
P24592	Insulin-like growth factor-binding protein 6	IGFBP6	2.058	1.8	1.548	1.2	0.098	0.016	0.555	0.256
P34932	Heat shock 70 kDa protein 4	HSPA4	2.136	3.9	2.037	2.7	0.070	0.240	0.127	0.897
Q06323	Proteasome activator complex subunit 1	PSME1	0.748	0.5	0.717	0.2	0.462	0.540	0.512	0.291

Q15435	Protein phosphatase 1 regulatory subunit 7	PPP1R7	1.777	2.9	1.678	2.1	0.735	0.540	0.748	0.126
Q6UY14	ADAMTS-like protein 4	ADAMTSL4	1.415	1.3	1.268	1.3	0.695	0.490	0.869	0.061
Q8WWA0	Intelectin-1	ITLN1	2.158	1.5	2.832	3.9	0.326	0.730	0.048	0.360
Q969H8	Myeloid-derived growth factor	C19orf10	2.555	4.3	1.293	5.2	0.234	0.220	0.925	0.034
Q99969	Retinoic acid receptor responder protein 2	RARR ES2	1.569	1.5	1.411	1.3	0.492	0.230	0.720	0.410
Q9BTY2	Plasma alpha-L-fucosidase	FUCA2	0.689	0.5	0.804	0.3	0.318	0.350	0.722	0.034
Q9NZK5	Adenosine deaminase 2	CECR1	0.723	0.6	0.654	0.5	0.401	0.310	0.344	0.464

4.3.7 Plasma proteome changes in preclinical AD– potential early biomarkers:

A retrospective analysis of baseline data allows us to identify potential early biomarkers of AD (i.e., ADW1/CTRLW1 ratios). I identified a total of 160 dysregulated proteins (Figure 4.9 and Table 4.12), including 59 upregulated and 101 downregulated proteins in preclinical AD (ADW1/CTRLW1). The volcano plot and heatmap of all AD preclinical DEPs are depicted in Figures 4.9A and 4.9B. The volcano plot shows the top 20 most upregulated and downregulated DEPs (Figure 4.9A) and include functions such as metabolism (AMY2B, BLVRB), regulation (PPP2R4, SERPINA1), cytoskeleton (KRTAP13-2), immunity (ITLN1, RNH1, CRP, CHIT1), and transport (ALB, TTR). In the volcano plot, the display shows the top ten downregulated DEPs in the preclinical AD group and manually identified roles using available literature and GO analysis. These DEPs were involved in antioxidant (PRDX5), apoptosis (BID), signalling/regulatory (DUSP3, RSU1, ARHGAP1, YWHAQ, PAM), cytoskeleton (ARPC2), metabolism (FABP1), and anticoagulant (ANXA5) (Figure 4.9A, Table 4.12) pathways.

A total of 15 DEPs were common to both incipient AD (ADW4/CTRLW4) and preclinical AD (ADW1/CTRLW1) (Figure 4.9C and Table 4.13). Of these, 9 DEPs were decreased in both clinical and preclinical AD, and included functions such as cytoskeleton/microtubule assembly (ARPC5, MAPRE2), signalling/regulation (PGK1), extracellular matrix (SPARC, VCL), apoptosis (VCP, ISOC1), protein folding (GANAB), and unknown (LGALSL). Two DEPs were increased in both preclinical and incipient AD: chitinase-3-like protein 1 (CHI3L1) and proteasome subunit beta type-2 (PSMB2). In addition, two DEPs were increased in preclinical AD but decreased in incipient AD: serotransferrin (TF) and Serine/threonine-protein phosphatase 2A (PPP2R4), and two proteins were decreased in preclinical AD and increased in incipient AD; fructose-bisphosphate aldolase B (ALDOB) and peptidyl glycine alpha-amidating monooxygenase (PAM). These 15 DEPs may be potential preclinical plasma biomarkers of early AD.

In preclinical AD, 59 upregulated proteins were associated with 93 biological processes, 18 cellular components, 2 molecular functions, and 22 KEGG & Reactome pathways with significant GO enrichments. These GO-term enrichments were complement activation, post-translational protein modification, inflammatory response, neutrophil degranulation, metabolism, proteasome, and immune system. The 101 downregulated proteins involved 143

biological processes, 42 cellular components, 26 molecular functions, and 35 KEGG. Most DEPs were found to be involved in one of the following categories: immune system, actin cytoskeleton and polymerization, response to unfolded proteins, protein binding, glycolysis/gluconeogenesis, signalling by Rho GTPases and haemostasis.

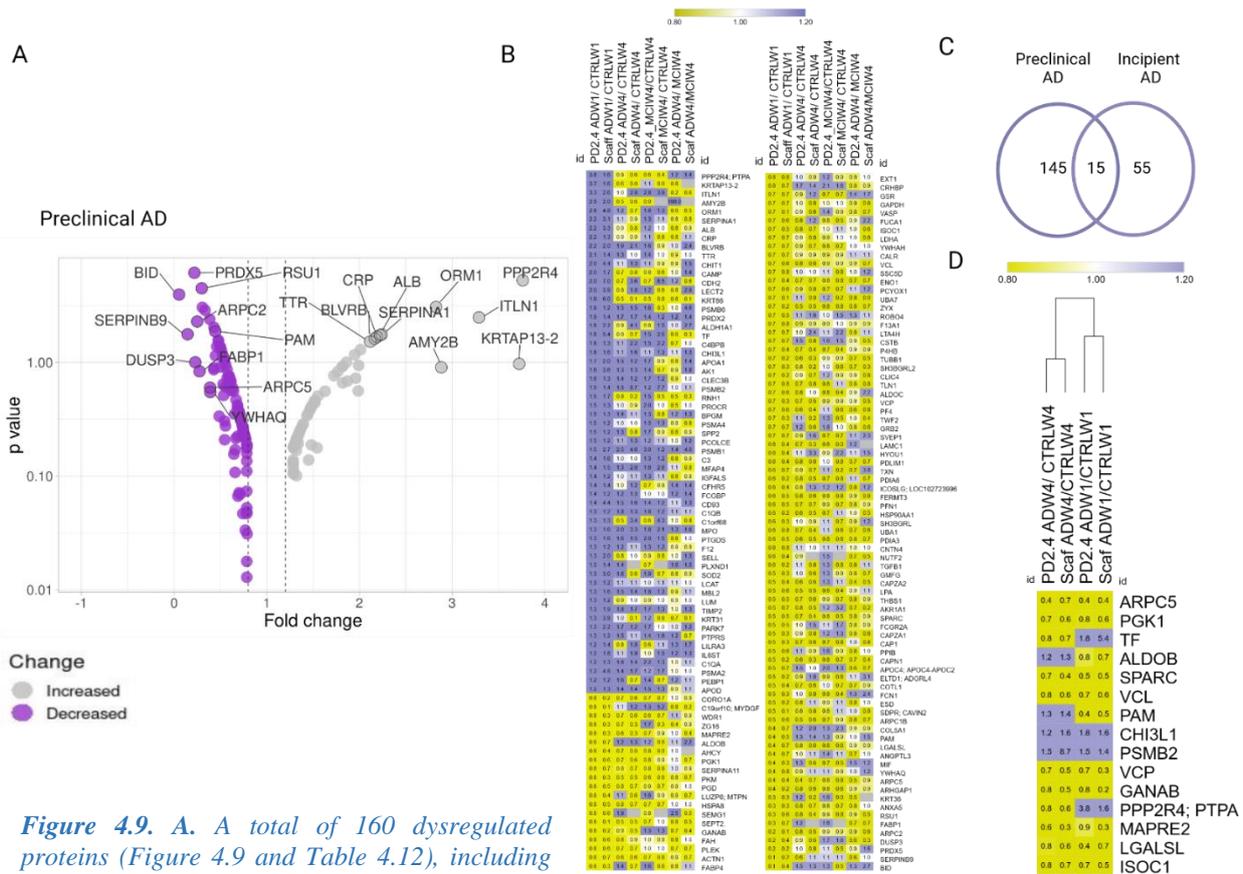


Figure 4.9. A. A total of 160 dysregulated proteins (Figure 4.9 and Table 4.12), including 59 upregulated and 101 downregulated proteins in preclinical AD. Volcano plots highlight the 20 DEPs with the highest fold change in preclinical AD (ADW1/CTRLW1) comparisons (I have highlighted only the top 20 proteins to avoid the overcrowding on volcano plots). Table 4.12 contained the detailed information of 160 DEPs of preclinical AD. B. Heatmap of 160 dysregulated proteins in preclinical AD presents all 160 DEPs in two panels to clearly show the fold change and protein names. C. I chose highly enriched GO from STRING, including upregulated and downregulated GO. D and E. Venn diagram showing the number of DEPs (15) which were common in preclinical AD (ADW1/CTRLW1) and cross-sectional AD (ADW4/CTRLW4), and these DEPs were further presented in the heatmap (Table 4.13).

Table 4.12. In preclinical AD, 160 proteins were dysregulated, including 59 upregulated and 101 downregulated proteins (ADW1/CTRLW1).

Accession	Description	Gene Symbol	PD2.4_Abundance Ratio: (ADW1) / (CTRLW1)	Scaffold fold change: (ADW1) / (CTRLW1)	PD2.4_Abundance Ratio Adj. P-Value: (ADW1) / (CTRLW1)	Scaffold P-Value: (ADW1) / (CTRLW1)
Q15257	Serine/threonine-protein phosphatase 2A activator	PPP2R4; PTPA	3.762	1.6	5.3528E-06	0.7
Q52LG2	Keratin-associated protein 13-2	KRTAP13-2	3.723	1.6	0.10676621	0.68
Q8WWA0	Intelectin-1	ITLN1	3.289	2.6	0.00325529	0.46
P19961	Alpha-amylase 2B	AMY2B	2.883	2	0.12436948	0.49
P02763	Alpha-1-acid glycoprotein 1	ORM1	2.826	4.8	0.00085114	0.12
P01009	Alpha-1-antitrypsin	SERPINA1	2.239	3.1	0.01708753	0.11
P02768	Serum albumin	ALB	2.222	3.3	0.01822968	0.046
P02741	C-reactive protein	CRP	2.188	1.3	0.02112758	0.56
P30043	Flavin reductase (NADPH)	BLVRB	2.164	2	0.02404333	0.14
P02766	Transthyretin	TTR	2.116	1.4	0.02986103	0.33
Q13231	Chitotriosidase-1	CHIT1	2.012	4.4	0.05050651	0.12
P49913	Cathelicidin antimicrobial peptide	CAMP	1.991	1.7	0.11444124	0.39
P19022	Cadherin-2	CDH2	1.986	7	0.27463361	0.078
O14960	Leukocyte cell-derived chemotaxin-2	LECT2	1.958	3.9	0.18776132	0.26
O43790	Keratin, type II cuticular Hb6	KRT86	1.948	6	0.06757085	0.21
P28072	Proteasome subunit beta type-6	PSMB6	1.87	1.2	0.20624972	0.85
P32119	Peroxiredoxin-2	PRDX2	1.861	1.4	0.09874915	0.29
P00352	Retinal dehydrogenase 1	ALDH1A1	1.819	2.2	0.21820511	0.47
P02787	Serotransferrin	TF	1.817	5.4	0.11955734	0.082
P20851	C4b-binding protein beta chain	C4BPB	1.816	1.3	0.11955734	0.35
P36222	Chitinase-3-like protein 1	CHI3L1	1.766	1.6	0.14768823	0.34
P02647	Apolipoprotein A-I	APOA1	1.724	2	0.17181951	0.26
P00568	Adenylate kinase isoenzyme 1	AK1	1.629	3.6	0.27631292	0.05

P05452	Tetranectin	CLEC3B	1.552	1.3	0.31088637	0.3
P49721	Proteasome subunit beta type-2	PSMB2	1.543	1.4	0.6646973	0.79
P13489	Ribonuclease inhibitor	RNH1	1.52	1.7	0.36025267	0.67
Q9UNN8	Endothelial protein C receptor	PROCR	1.515	1.3	0.34582552	0.39
P07738	Bisphosphoglycerate mutase	BPGM	1.489	1.3	0.3798756	0.5
P25789	Proteasome subunit alpha type-4	PSMA4	1.481	1.2	0.64828595	0.83
Q13103	Secreted phosphoprotein 24	SPP2	1.474	1.2	0.40069352	0.61
Q15113	Procollagen C-endopeptidase enhancer 1	PCOLCE	1.465	1.2	0.41607405	0.32
P20618	Proteasome subunit beta type-1	PSMB1	1.452	2.7	0.72685903	0.36
P01024	Complement C3	C3	1.444	1.6	0.44300254	0.26
P55083	Microfibril-associated glycoprotein 4	MFAP4	1.433	1.5	0.45788588	0.6
P35858	Insulin-like growth factor-binding protein complex acid labile subunit	IGFALS	1.393	1.2	0.51641443	0.32
Q9BXR6	Complement factor H-related protein 5	CFHR5	1.383	1.4	0.52794083	0.18
Q9Y6R7	IgGfc-binding protein	FCGBP	1.363	1.2	0.55458471	0.51
Q9NPY3	Complement component C1q receptor	CD93	1.353	4.4	0.56974874	0.057
P02746	Complement C1q subcomponent subunit B	C1QB	1.341	1.2	0.58837601	0.27
Q5T750	Skin-specific protein 32	C1orf68	1.34	1.3	0.64017955	0.75
P05164	Myeloperoxidase	MPO	1.335	1.6	0.59544863	0.54
P41222	Prostaglandin-H2 D-isomerase	PTGDS	1.333	1.6	0.59861347	0.097
P00748	Coagulation factor XII	F12	1.325	1.2	0.6085159	0.33
P14151	L-selectin	SELL	1.324	2	0.61115963	0.022
Q9Y4D7	Plexin-D1	PLXND1	1.32	1.4	0.79266266	0.76
P04179	Superoxide dismutase [Mn], mitochondrial	SOD2	1.319	3	0.67863625	0.32

P04180	Phosphatidylcholine-sterol acyltransferase	LCAT	1.317	1.2	0.61806649	0.39
P11226	Mannose-binding protein C	MBL2	1.317	1.6	0.61806649	0.16
P51884	Lumican	LUM	1.306	1.2	0.63839034	0.014
P16035	Metalloproteinase inhibitor 2	TIMP2	1.292	1.9	0.65371813	0.1
Q15323	Keratin, type I cuticular Ha1	KRT31	1.289	3.9	0.65687155	0.37
Q99497	Protein/nucleic acid deglycase DJ-1	PARK7	1.287	2.2	0.65910847	0.072
Q13332	Receptor-type tyrosine-protein phosphatase S	PTPRS	1.279	1.2	0.66468896	0.71
Q8N6C8	Leukocyte immunoglobulin-like receptor subfamily A member 3	LILRA3	1.241	1.4	0.72685903	0.56
P40189	Interleukin-6 receptor subunit beta	IL6ST	1.228	1.8	0.73859717	0.31
P02745	Complement C1q subcomponent subunit A	C1QA	1.222	1.3	0.74809978	0.37
P25787	Proteasome subunit alpha type-2	PSMA2	1.212	4.8	0.7844346	0.024
P30086	Phosphatidylethanolamine-binding protein 1	PEBP1	1.212	1.2	0.7661659	0.7
P05090	Apolipoprotein D	APOD	1.2	1.3	0.78324037	0.33
P31146	Coronin-1A	CORO1A	0.896	0.2	0.88576186	0.21
Q969H8	Myeloid-derived growth factor	C19orf10; MYDGF	0.888	0.1	0.9707224	0.19
O75083	WD repeat-containing protein 1	WDR1	0.876	0.3	0.84373271	0.28
O60844	Zymogen granule membrane protein 16	ZG16	0.861	0.3	0.95987143	0.15
Q15555	Microtubule-associated protein RP/EB family member 2	MAPRE2	0.856	0.3	0.89783133	0.19
P05062	Fructose-bisphosphate aldolase B	ALDOB	0.846	0.7	0.7751356	0.21
P23526	Adenosylhomocysteinase	AHCY	0.843	0.4	0.89449043	0.56
P00558	Phosphoglycerate kinase 1	PGK1	0.808	0.6	0.69374295	0.34
Q86U17	Serpin A11	SERPINA11	0.798	0.7	0.6649673	0.35
P14618	Pyruvate kinase PKM	PKM	0.795	0.3	0.65963294	0.18

P52209	6-phosphogluconate dehydrogenase, decarboxylating	PGD	0.793	0.5	0.65910847	0.37
P58546	Myotrophin	LUZP6; MTPN	0.79	0.4	0.73151142	0.4
P11142	Heat shock cognate 71 kDa protein	HSPA8	0.789	0.5	0.65371813	0.39
P04279	Semenogelin-1	SEMG1	0.788	0.6	0.93101197	0.74
Q15019	Septin-2	SEPT2	0.786	0.1	0.64828595	0.18
Q14697	Neutral alpha-glucosidase AB	GANAB	0.784	0.2	0.64727023	0.19
P16930	Fumarylacetoacetase	FAH	0.772	0.6	0.61534114	0.078
P08567	Pleckstrin	PLEK	0.77	0.6	0.61002491	0.47
P12814	Alpha-actinin-1	ACTN1	0.768	0.7	0.6065492	0.53
P15090	Fatty acid-binding protein, adipocyte	FABP4	0.768	0.3	0.92576353	0.36
Q16394	Exostosin-1	EXT1	0.765	0.8	0.59861347	0.58
P24387	Corticotropin-releasing factor-binding protein	CRHBP	0.759	0.7	0.58609407	0.33
P00390	Glutathione reductase, mitochondrial	GSR	0.749	0.7	0.56068472	0.19
P04406	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	0.746	0.7	0.55085056	0.27
P50552	Vasodilator-stimulated phosphoprotein	VASP	0.746	0.1	0.55072217	0.16
P04066	Tissue alpha-L-fucosidase	FUCA1	0.744	0.6	0.54571062	0.35
Q96CN7	Isochorismatase domain-containing protein 1	ISOC1	0.742	0.5	0.89783133	0.62
P00338	L-lactate dehydrogenase A chain	LDHA	0.74	0.7	0.53310518	0.26
Q04917	14-3-3 protein eta	YWHAH	0.74	0.7	0.58327636	0.66
P27797	Calreticulin	CALR	0.739	0.7	0.53153453	0.67
P18206	Vinculin	VCL	0.737	0.6	0.52799201	0.37
A1L4H1	Soluble scavenger receptor cysteine-rich domain-containing protein SSC5D	SSC5D	0.733	0.8	0.51821418	0.76

P06733	Alpha-enolase	ENO1	0.724	0.4	0.49535944	0.14
Q9UHG3	Prenylcysteine oxidase 1	PCYOX1	0.724	0.6	0.49535944	0.041
P41226	Ubiquitin-like modifier-activating enzyme 7	UBA7	0.712	0.1	0.85232222	0.36
Q15942	Zyxin	ZYX	0.709	0.2	0.45060653	0.13
Q8WZ75	Roundabout homolog 4	ROBO4	0.706	0.5	0.44228768	0.17
P00488	Coagulation factor XIII A chain	F13A1	0.705	0.8	0.43958717	0.32
P09960	Leukotriene A-4 hydrolase	LTA4H	0.705	0.7	0.43999829	0.57
P04080	Cystatin-B	CSTB	0.701	0.7	0.84859405	0.7
P07237	Protein disulfide-isomerase	P4HB	0.698	0.4	0.42213929	0.29
Q9H4B7	Tubulin beta-1 chain	TUBB1	0.698	0.2	0.42170309	0.14
Q9UJC5	SH3 domain-binding glutamic acid-rich-like protein 2	SH3BGRL2	0.689	0.3	0.85786557	0.47
Q9Y696	Chloride intracellular channel protein 4	CLIC4	0.687	0.2	0.59544863	0.24
Q9Y490	Talin-1	TLN1	0.685	0.8	0.38029616	0.76
P09972	Fructose-bisphosphate aldolase C	ALDOC	0.683	0.6	0.39240579	0.34
P55072	Transitional endoplasmic reticulum ATPase	VCP	0.669	0.3	0.33885529	0.3
P02776	Platelet factor 4	PF4	0.666	0.6	0.61002491	0.67
Q6IBS0	Twinfilin-2	TWF2	0.659	0.3	0.67525933	0.39
P62993	Growth factor receptor-bound protein 2	GRB2	0.658	0.7	0.78009507	0.81
Q4LDE5	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	SVEP1	0.653	0.7	0.44508596	0.73
P11047	Laminin subunit gamma-1	LAMC1	0.645	0.4	0.69781554	0.34
Q9Y4L1	Hypoxia up-regulated protein 1	HYOU1	0.641	0.4	0.27454849	0.34
O00151	PDZ and LIM domain protein 1	PDLIM1	0.632	0.4	0.25357436	0.23
P10599	Thioredoxin	TXN	0.619	0.7	0.27268199	0.54
Q15084	Protein disulfide-isomerase A6	PDIA6	0.613	0.2	0.20888629	0.16

O75144	ICOS ligand	ICOSLG; LOC10272399 6	0.61	0.4	0.20624972	0.26
Q86UX7	Fermitin family homolog 3	FERMT3	0.603	0.5	0.19186621	0.3
P07737	Profilin-1	PFN1	0.594	0.5	0.17733864	0.32
P07900	Heat shock protein HSP 90-alpha	HSP90AA1	0.59	0.2	0.17068273	0.24
O75368	SH3 domain-binding glutamic acid-rich-like protein	SH3BGRL	0.589	0.3	0.23173268	0.27
P22314	Ubiquitin-like modifier-activating enzyme 1	UBA1	0.572	0.8	0.18603507	0.71
P30101	Protein disulfide-isomerase A3	PDIA3	0.571	0.5	0.13526359	0.26
Q8IYW2	Contactin-4	CNTN4	0.556	0.8	0.11187858	0.75
P61970	Nuclear transport factor 2	NUTF2	0.554	0.4	0.529576	0.55
P01137	Transforming growth factor beta-1	TGFB1	0.551	0.2	0.24499433	0.23
O60234	Glia maturation factor gamma	GMFG	0.545	0.3	0.49562623	0.38
P47755	F-actin-capping protein subunit alpha-2	CAPZA2	0.545	0.4	0.61806649	0.25
P08519	Apolipoprotein(a)	LPA	0.543	0.6	0.0909234	0.37
P07996	Thrombospondin-1	THBS1	0.536	0.5	0.08166281	0.24
P14550	Alcohol dehydrogenase [NADP(+)]	AKR1A1	0.535	0.7	0.30687761	0.82
P09486	SPARC	SPARC	0.529	0.5	0.07392471	0.36
P12318	Low affinity immunoglobulin gamma Fc region receptor II-a	FCGR2A	0.528	0.5	0.07382711	0.43
P52907	F-actin-capping protein subunit alpha-1	CAPZA1	0.528	0.3	0.2278709	0.17
Q01518	Adenylyl cyclase-associated protein 1	CAP1	0.51	0.3	0.05354235	0.24
P23284	Peptidyl-prolyl cis-trans isomerase B	PPIB	0.502	0.6	0.04580394	0.31
P07384	Calpain-1 catalytic subunit	CAPN1	0.501	0.2	0.04461278	0.21

P55056	Apolipoprotein C-IV	APOC4; APOC4- APOC2	0.501	0.7	0.04469192	0.49
Q9HBW9	Adhesion G protein-coupled receptor L4	ELTD1; ADGRL4	0.498	0.2	0.12301082	0.23
Q14019	Coactosin-like protein	COTL1	0.497	0.4	0.04187312	0.24
O00602	Ficolin-1	FCN1	0.486	0.3	0.4606749	0.28
P10768	S-formylglutathione hydrolase	ESD	0.48	0.2	0.02914401	0.078
O95810	Caveolae-associated protein 2	SDPR; CAVIN2	0.476	0.06	0.02663948	0.14
O15143	Actin-related protein 2/3 complex subunit 1B	ARPC1B	0.461	0.6	0.04146083	0.56
P20908	Collagen alpha-1(V) chain	COL5A1	0.447	0.7	0.07131572	0.54
P19021	Peptidyl-glycine alpha-amidating monooxygenase	PAM	0.444	0.5	0.01319158	0.38
Q3ZCW2	Galectin-related protein	LGALS1	0.431	0.7	0.12764282	0.77
Q9Y5C1	Angiopoietin-related protein 3	ANGPTL3	0.43	0.7	0.00921395	0.56
P14174	Macrophage migration inhibitory factor	MIF	0.428	0.3	0.13684294	0.35
P27348	14-3-3 protein theta	YWHAQ	0.39	0.8	0.28245865	0.73
O15511	Actin-related protein 2/3 complex subunit 5	ARPC5	0.388	0.4	0.25084048	0.46
Q07960	Rho GTPase-activating protein 1	ARHGAP1	0.384	0.2	0.00399729	0.27
O76013	Keratin, type I cuticular Ha6	KRT36	0.334	0.3	0.00133438	0.36
P08758	Annexin A5	ANXA5	0.305	0.3	0.00088785	0.17
Q15404	Ras suppressor protein 1	RSU1	0.298	0.6	3.2609E-05	0.54
P07148	Fatty acid-binding protein, liver	FABP1	0.274	0.7	0.14598386	0.7
O15144	Actin-related protein 2/3 complex subunit 2	ARPC2	0.249	0.1	0.00514409	0.09
P51452	Dual specificity protein phosphatase 3	DUSP3	0.229	0.4	0.09937377	0.53
P30044	Peroxiredoxin-5, mitochondrial	PRDX5	0.219	0.2	7.1634E-07	0.22

P50453	Serpin B9	SERPINB9	0.145	0.2	0.01696295	0.36
P55957	BH3-interacting domain death agonist	BID	0.054	0.4	0.00011138	0.53

Table 4.13. A total of 15 DEPs that were found to be commonly differentially expressed in age-matched clinical cross-sectional incipient AD (ADW4/CTRLW4) from Table 4.7 and preclinical AD (ADW1/CTRLW1) protein lists Table 4.12.

Accession	Description	Gene Symbol	PD2.4_Abundance Ratio: (ADW4) / (CTRLW4)	Scaffold fold change (ADW4) / (CTRLW4)	PD2.4_Abundance Ratio: (ADW1) / (CTRLW1)	Scaffold fold change: (ADW1) / (CTRLW1)	PD2.4_Abundance Ratio Adj. P-Value: (ADW4) / (CTRLW4)	Scaffold P-Value: (ADW4) / (CTRLW4)	PD2.4_Abundance Ratio Adj. P-Value: (ADW1) / (CTRLW1)	Scaffold P-Value: (ADW1) / (CTRLW1)
O15511	Actin-related protein 2/3 complex subunit 5	ARPC5	0.374	0.7	0.388	0.4	0.0459	0.44	0.2508	0.46
P00558	Phosphoglycerate kinase 1	PGK1	0.716	0.6	0.808	0.6	0.4812	0.068	0.6937	0.34
P02787	Serotransferrin	TF	0.869	0.7	1.817	5.4	0.8416	0.54	0.1196	0.082
P05062	Fructose-bisphosphate aldolase B	ALDOB	1.221	1.3	0.846	0.7	0.7317	0.45	0.7751	0.21
P09486	SPARC	SPARC	0.675	0.4	0.529	0.5	0.3351	0.0089	0.0739	0.36
P18206	Vinculin	VCL	0.779	0.6	0.737	0.6	0.6553	0.024	0.528	0.37
P19021	Peptidyl-glycine alpha-amidating monooxygenase	PAM	1.27	1.4	0.444	0.5	0.6553	0.59	0.0132	0.38
P36222	Chitinase-3-like protein 1	CHI3L1	1.142	1.6	1.766	1.6	0.8396	0.18	0.1477	0.34
P49721	Proteasome subunit beta type-2	PSMB2	1.503	8.7	1.543	1.4	0.6529	0.083	0.6647	0.79
P55072	Transitional endoplasmic reticulum ATPase	VCP	0.704	0.5	0.669	0.3	0.4356	0.14	0.3389	0.3
Q14697	Neutral alpha-glucosidase AB	GANAB	0.857	0.5	0.784	0.2	0.8194	0.17	0.6473	0.19

Q15257	Serine/threonine-protein phosphatase 2A activator	PPP2R4; PTPA	0.863	0.6	3.762	1.6	0.854	0.4	5E-06	0.7
Q15555	Microtubule-associated protein RP/EB family member 2	MAPRE2	0.598	0.3	0.856	0.3	0.3215	0.0086	0.8978	0.19
Q3ZCW2	Galectin-related protein	LGALS L	0.82	0.6	0.431	0.7	0.7595	0.15	0.1276	0.77
Q96CN7	Isochorismatase domain-containing protein 1	ISOC1	0.753	0.7	0.742	0.5	0.83	0.56	0.8978	0.62

4.4 Discussion:

This study has shown that a rich abundance of age and disease-related proteomic changes are detectable in plasma, based on retrospective analysis of longitudinal data and cross-sectional analyses of clinically diagnosed cases. In combination with a method which provides the depth of plasma proteome coverage (Chapter 3)³⁸, the study design has addressed the following questions: (1) differences in plasma proteomic profiles between normal ageing and ageing with progression to cognitive decline (MCI) or AD; (2) cross-sectional analysis of baseline data, when all subjects were clinically identified as cognitively normal, provides insight into the preclinical changes which precede subsequent progression to AD, and potentially provide early biomarkers; and (3) comparison of plasma at the point of progression to the clinically diagnosed onset of cognitive decline or AD, can provide potential plasma biomarkers to facilitate clinical diagnosis. I perceive two major obstacles in identifying plasma protein biomarkers for the common age-related neurodegenerative diseases: (1) the restricted current level of information regarding the plasma proteome longitudinal changes in normal vs diseased individuals, and (2) the even more limited knowledge of preclinical AD plasma proteome. I have begun to address both of these deficiencies in this work.

4.4.1 Plasma proteome changes in ageing using a longitudinal analysis:

Ageing is the primary factor associated with organ function decline, including age-related cognitive decline, and is a major risk factor for neurodegenerative diseases such as AD²⁶⁸. Consequently, it is common to use age-matched controls to study such disorders and diseases. However, the extent of change in the ageing plasma proteome, irrespective of disease, is less clear. Here I identified 71 proteins that were dysregulated during normal ageing, with the majority being increased. These proteins were identified in all three of our longitudinal groups (normal controls, MCI and AD, Table 4.2) and with a similar fold-change direction (Table 4.3). The Hippo signalling pathway was particularly enriched with ageing (Figure 4.10). This signalling pathway included DEPs of the 14-3-3 protein family (YWHAZ, YWHAH, YWHAE, YWHAB, YWHAQ, YWHAB) and actin Gamma 1 (ACTG1), which were all upregulated (Table 4.3). Recent evidence suggests that the Hippo signalling pathway is involved in neuroinflammation, neuronal cell differentiation, and neuronal death¹⁶⁹. The 14-3-3 protein family is highly expressed in the brain and influences many aspects of brain function

through interactions with a diverse set of binding partners, including neural signalling, neuronal development, and neuroprotection¹⁷⁰ well-studied protein family in AD CSF^{171,172}. Our longitudinal analysis shows that altered plasma expression of the 14-3-3 protein family is an age-related change, being observed in all three longitudinal analysis groups (cognitively normal controls, MCI and AD), so it may have functional implications for progression to MCI and/or AD since ageing is the major risk factor for these conditions. However, as the hippo family members are not unique to AD (Table 4.3, Figure 4.4), they are less likely to be valuable biomarkers.

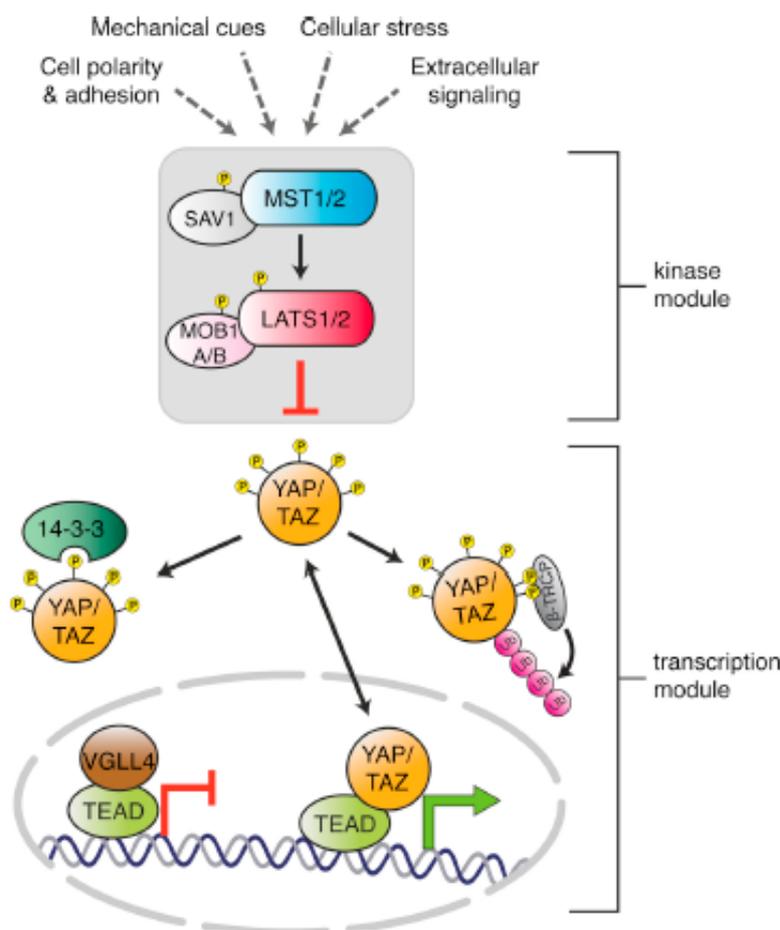


Figure 4.10: Schematic model of the core Hippo pathway in mammals. Several upstream stimuli transduce signals that stimulate the MST1/2 kinases with the scaffolding protein SAV1 to phosphorylate the LATS1/2 kinases and their regulatory subunit MOB1A/B. The activated LATS1/2-MOB1A/B complex phosphorylates YAP/TAZ, resulting in cytoplasmic retention by 14-3-3 protein binding and β -TRCP-mediated polyubiquitination and proteasomal degradation of YAP/TAZ. Unphosphorylated YAP/TAZ shuttle between cytoplasm and nucleus. The nucleus interacts mainly with TEAD transcription factors to induce target gene transcription. In the absence of nuclear YAP/TAZ, TEAD interacts with VGLL4, thereby repressing target gene transcription. This figure was adopted from Dana Elster and Björn von Eyss 2020²⁶⁹.

Damaged and misfolded proteins accumulate during the ageing process, affecting cell function and tissue homeostasis. Cellular clearance processes such as the proteasome are a critical component of the proteostasis network, involved in the degradation and recycling of damaged proteins. Proteasome activity declines with age, and dysfunctional proteasomes are related to late-onset diseases²⁷⁰. I identified five dysregulated proteasome members, all of which were

upregulated in plasma: PSMA4, PSME2, PSMA2, PSMA5, and PSMB8, suggesting that intracellular protein turnover is compromised with ageing. Other protein families of which multiple members were identified in our longitudinal ageing groups include actin-related protein (6 DEPs), chloride intracellular channel protein (2 DEPs), glutathione S-transferase (2 DEPs), L-lactate dehydrogenase (2 DEPs), peptidyl-prolyl cis-trans isomerase (2 DEPs), protein S100 (2 DEPs), and rho GDP-dissociation inhibitor (2 DEPs) (Table 4.2 and Table 4.3). Since most of these are intracellular proteins, their presence in plasma is likely a reflection of cellular senescence, increased fragility and cell death with ageing. Therefore, the functions they sub-serve are likely compromised with ageing and may predispose to disease progression. However, since they were also present in the cognitively normal ageing group, these changes are insufficient on their own to explain progression to cognitive disorder or neurodegenerative disease. Changes to many of these proteins have previously been attributed to associations with either MCI and/or AD²⁷¹⁻²⁷³. The current work demonstrates the need for particular care in age matching in case-control studies, especially biomarker studies.

A variety of other age-related protein changes were observed in common across all three longitudinal analysis groups (normal controls, MCI and AD), including the HIF-1 signalling pathway (Table 4.3) and several proteins abundant in the CNS (NCAM1, YWHA family, PKM, NME2, MAPRE1) indicating that age-related changes within the CNS can be detected in plasma, with techniques which allow sufficient depth of proteome coverage. Previously, studies showed that HIF-1 generates a deficit in mitochondrial biogenesis during the ageing process impairing energy-dependent cellular functions such as cell and tissue repair²⁷⁴. I have identified a list of markers such as TIMP1, GAPDH, ENO1, PGK1, LDHB, and LDHA that can aid in the understanding of mitochondrial dysregulation in ageing (Table 4.3). In addition, a HIF-1 signalling pathway is known to both promote and limit longevity via pathways that are mechanistically distinct using hypoxic response transcription factor HIF-1^{275,276}. Validation of these DEPs in large sample size cohorts might improve our understanding of human ageing. Such broad-ranging pathway changes in ageing may also help explain why ageing is the single major risk factor for a wide variety of diseases, including age-related neurodegenerative diseases such as dementia. This wide range of pathways impacted by the ageing process likely overlaps with many disease processes, making ageing an accelerant if not a causative risk factor for cognitive decline and/or disease.

4.4.2 Changes in the plasma proteome associated with progression to incipient AD or MCI over 6 years:

To identify MCI and AD specific changes in the longitudinal analysis groups, I filtered the DEPs list for age-independent protein changes, which were unique to either AD or MCI, but not similarly changed in cognitively normal ageing (Table 4.4 and 4.5). A characteristic of AD patients that brain imaging techniques can detect is impaired glucose uptake in brain regions with neuritic plaques²⁷⁷⁻²⁷⁹. Numerous dementia probable AD-specific DEPs involved in metabolism (e.g., APOD, ALDOB, MAN2A2, GPX3, HPRT1, ALDH1A1, AMY1A, MGAT1, and IGFBP5) were elevated in plasma in our investigation, reflecting impairment of the cellular metabolic processes in which these proteins partake.

APOD is a glial-expressed lipid transport protein of the lipocalin family that has been shown to protect against oxidative stress²⁸⁰. Our longitudinal data show that increased APOD is observed in MCI and AD plasma, but not in cognitively normal ageing (Tables 4.4 and 4.5). This observation is consistent with other published work, which shows elevated APOD in AD, Parkinson's disease, Schizophrenia, Stroke and Bipolar disorder^{280,281,282}. Increased plasma IGFBP5 also appears to be related to cognitive decline since older adults with depression lose cognitive abilities faster when they have higher IGFBP-5 levels²⁸³. HPRT1 was recently identified as one of the most strongly validated metabolic proteins, exhibiting a substantial increase in AD CSF cohorts, demonstrating a direct link between energy production and synaptic signalling at the neuronal membrane^{171,284}. Another metabolic protein identified is ALDH1A1, a multifunctional enzyme with dehydrogenase, esterase, and antioxidant activities and critical for normal brain homeostasis, which was upregulated in AD downregulated in MCI in our data. A recent study shows neurons may upregulate ALDH1A1 activity to compensate for oxidative stress-induced damage in the brain²⁸⁵. I am proposing metabolic abnormalities, which can be identified in plasma, as a critical component of longitudinal AD aetiology, a better understanding of which might provide novel metabolic targets for therapeutic development.

In addition to metabolism, a large group of proteins was associated with homeostasis in AD and MCI. Homeostasis related proteins were upregulated in MCI but downregulated in AD. A recent study suggested that firing instability and poor synaptic plasticity during the early stages of AD initiate a vicious loop that results in integrated homeostatic network (IHN) dysregulation²⁸⁶. According to this idea, the collapse of the IHN is the primary factor driving the transition from

early memory deficits to neurodegeneration²⁸⁷. Homeostatic proteins which were downregulated in AD are COL1A1, SELL, ENDOD1, TF, SERPINA1. Decreased level of TF in AD plasma and brain samples has been reported previously^{288,289}. Consistent with these early reports, I found significant downregulation of TF during longitudinal progression to AD in our study.

In MCI, several unique homeostasis markers were differentially expressed, including upregulation of RAB27B and PPBP, which may participate in the pathology underlying MCI. RAB27a and RAB27b are involved in the docking of MVE at the plasma membrane²⁹⁰. Previous studies have found that the upregulated expression of RAB27 correlated with antemortem indicators of cognitive deterioration in MCI and AD brains²⁹⁰⁻²⁹². In our data, RAB27B was elevated in MCI but did not change in AD, which may imply endosomal dysfunction as an early change detected in MCI, which may contribute to progression to AD in later stages of life. Alternatively, it may also be a change specific to MCI, and studies of longer duration may help decipher what changes are associated with stable MCI vs progression to AD.

It is believed that the extracellular matrix contains collagens, which are essential in axonal guidance, synaptogenesis, cell adhesion, the formation of brain architecture and neural maturation²⁹³⁻²⁹⁵. A gene from the collagen family, *Col25a1*, was overexpressed in neurons of transgenic mice leading to AD-like brain pathology²⁹⁶. In our data, COL1A1 was upregulated in longitudinal MCI progression but downregulated in AD. Such differences may point to mechanisms that help limit the level of impairment and avoid progression to greater levels of cognitive impairment.

Moreover, SERPINA1 is emerging as a key neuroinflammation modulator²⁹⁷, also reported being released from the brain tissue to the CSF²⁹⁸. Higher CSF levels of SERPINA1 have been linked to the clinical diagnosis of AD²⁹⁹. Here I found that SERPINA1 was upregulated in MCI and also in preclinical AD but downregulated in clinical AD (Figure 4.9B), suggesting that it may be an early marker of synaptic loss particularly evident in plasma at preclinical stages of dementia, at a time when much damage is in active progress, and plateauing/declining in parallel with the onset of clinical symptoms.

4.4.3 Proteomic changes in clinically diagnosed AD and MCI relative to their age-matched cognitively normal controls (potential clinical diagnostic markers):

A total of 70 and 89 DEPs were identified in a cross-sectional analysis of incipient AD (ADW4/CRTLW4) and incipient MCI (MCIW4/CRTLW4), respectively, indicating that a potentially rich biomarker signature for AD and MCI is available in plasma.

Of the 70 cross-sectional AD-associated DEPs, 15 were also identified in longitudinal AD analysis (Table 4.10), of which 11 had a consistent direction of expression change. These may be the most robust biomarker candidates as they change consistently in longitudinal and cross-sectionally dementia, probable relative to their age-matched controls (Table 4.10). Of these, 10 DEPs (TF, VCP, PSMB2, PA2G4, PAM, MAN2A2, TNXB, FAM3C, ALDOB, and QDPR) were enriched with extracellular exosome GO terms in the STRING analysis. At an early stage of AD, a rise in the protein levels of total and phosphorylated tau in exosomes has been found in the CSF^{300,301}. Another finding implies that exosomes may be the primary mechanism controlling the spread of A β and tau³⁰². Our findings are consistent with the published literature, which indicates that exosome dysregulation is a key event in AD patients compared to their age-matched healthy controls³⁰¹. In addition to homeostasis and metabolic disruption, neutrophil degranulation, protein binding, and transport were the most enriched pathways in incipient AD-related DEPs in cross-sectional analysis. Neutrophil activation and accompanying oxidative stress have been linked to AD pathogenesis³⁰³. It is noteworthy that our study identified brain-derived proteins such as MAN2A2, PAM, TF, QDPR, FAM3C, which have previously been reported to be dysregulated in AD CSF and brain³⁰⁴.

There were 18 DEPs common to both MCI and AD, which may be considered potential shared biomarkers of cognitive change, including FAM3C, TF, CLIC1, CHI3L1 PRDX4 and others (Figure 4.7D and E), and possibly reflecting the underlying disease process. FAM3C is an interleukin-like protein (also called ILEI) with a proposed role as a metabolic regulator³⁰⁵. FAM3C ameliorates A β pathology by reducing A β levels³⁰⁶, has been suggested as a surrogate biomarker of A β in the brain³⁰⁷ and FAM3C levels are lower in the AD brain³⁰⁸. Its normal expression level is exceptionally high in the gut, thyroid and brain (<https://www.ncbi.nlm.nih.gov/gene/10447>). Previous work has reported lower levels of CSF FAM3C in AD and MCI groups which suggested this may result in a build-up of A β in the brain and eventual development of AD³⁰⁸. I observed a higher level of FAM3C in plasma samples of MCI and dementia; probable AD compared to respective age-matched controls suggests either loss from the CNS or a homeostatic/compensatory increase in response to loss in an organ system/s.

The protein CHI3L1 (also known as YKL-40) is secreted by activated microglia and reactive astrocytes³⁰⁹ and is thought to have a role in inflammation and tissue remodelling, particularly angiogenesis³¹⁰. In the current work, CHI3L1 was increased in both MCI and AD plasma (Figure 4.7E), consistent with reported observations of higher CHI3L1 levels in AD than in healthy controls or MCI patients³¹¹, and also of other neurodegenerative diseases³¹². In addition, several studies have reported that CHI3L1, an astrocyte-derived protein, is increased in AD CSF and has been suggested to be a marker for progression from MCI to AD^{21,313}.

Serotransferrin (TF) decreased in incipient dementia and probable AD while increased in MCI²⁸⁹. Transferrin (Tf) is an important iron-binding protein that is thought to have a critical function in iron ion (Fe) absorption via the transferrin receptor (TfR). Elevated Fe levels in AD brains have been reported and linked to amyloid plaque formation³¹⁴.

4.4.4 Potential early disease markers; preclinical changes in AD

Preclinical AD, defined as a stage of neurodegeneration occurring before the onset of clinical symptoms, is likely to be the more effective time point to apply potentially disease-modifying interventions in AD^{142,143}. A total of 160 DEPs of preclinical AD (ADW1/CTRLW1) were identified (Figure 4.9), which were a considerably larger number than the 71 DEPs identified in incipient AD (Figure 4.8). The considerably larger pool of preclinical AD-associated DEPs may in part be evidence of pathology in progress, in addition to providing several putative early biomarkers. Reasoning that the most robust biomarkers may be those that continue to be observed with clinical disease onset, 15 DEPs were shared with clinical AD (ADW4/CTRLW4) (Figure 4.9C, 4.9D and Table 4.13). Furthermore, of these 15 DEPs, 8 were unique to preclinical and clinical AD but not identified in clinical MCI (MCIW4/CTRLW4). These 8 proteins were PGK1, ALDOB, VCL, PAM, PSMB2, VCP, MAPRE2, and ISOC1, and may be specific to AD-related pathology, rather than just associated with cognitive decline *per se*. Interestingly, glycolysis and gluconeogenesis presented as top GO terms with significant enrichment in preclinical and clinical AD plasma. This concurs with the presence of three glycolytic proteins from our 8-protein signature: PGK1, VCP and ALDOB. Numerous studies have demonstrated dysregulation of glucose metabolism in the brain, which has long been recognised as an apparent anomaly that commences during the preclinical stage of AD^{315,316,317} and remains a feature with incipient AD. Apart from the well-known CSF AD biomarker (CHI3L1), I propose a list of novel markers, including PSMB2, PAM, ALDOB, TF, MAPRE2, VCP, which may be potential preclinical biomarkers for the identification of AD, being dysregulated in all three AD comparison groups, i.e., longitudinal (ADW4/ADW1), incipient AD (ADW4/CTRLW4) and preclinical AD

(ADW1/CTRLW1). That they are all representative of different aspects of AD pathology (PSMB2, proteasomal turnover of dysfunctional proteins; PAM, signalling peptide synthesis, ALDOB, glycolysis and gluconeogenesis metabolic pathways; TF, iron-binding and transport; MAPRE2, a microtubule-associated protein with a possible role in signal transduction; VCP, segregation of proteins for degradation by the proteasome) may offer a specificity advantage, since AD is a complex multifactorial disease with dysfunction of multiple cellular pathways. It is of note that of these 6 proteins, the only DEP with a consistent fold-change (increased) across all three groups is PSMB2.

4.4.5 Mechanisms of AD

MCI is often considered a risk factor and/or prodromal stage of AD³¹⁸, so it was of note that in the comparisons of AD and MCI, only approximately 18% (19 proteins) and 13% (18 proteins) of DEPs were common to both AD and MCI in the longitudinal and cross-sectional analyses. In this context, it is of interest that by far most DEPs identified in AD and MCI are specific to each condition rather than shared.

Another prevalent hypothesis is that dysfunction of the cytoskeleton and microtubule system may contribute to AD pathogenesis^{319,320}. MAPRE2, a microtubule-associated protein, and VCL, a membrane-cytoskeletal protein, are involved in microtubule polymerization, cell-cell and cell-matrix junctions. A recent study suggested that MAPRE2 is involved in cellular migration of cranial neural crest cells, among others, via its involvement in focal adhesion dynamics³²¹, although no direct association between MAPRE2 and AD progression has been established previously.

Dysregulation of phosphorylation in AD is commonly observed³²², so it is interesting that several proteins in our preclinical biomarker list are directly or indirectly involved in phosphorylation (PGK1, GANAB, PPP2R4). Several studies have reported that GANAB and PPP2R4 are dysregulated in AD CSF and brain^{323,324}. Most of the brain Ser/Thr phosphatase activity involves PP2A family enzymes. The dysfunction of PPP2R4 has been linked to tau hyperphosphorylation, amyloidogenesis and synaptic deficits that are pathological hallmarks of AD^{324,325}. Furthermore, SPARC and ARPC5 are proteins involved in regulating cell-cell interactions, actin polymerisation and neural plasticity, respectively. It has been reported that chronic stress significantly increased the level of ARPC5 in the hippocampus, implying that chronic stress-induced alterations in hippocampal proteins are related to synaptic plasticity³²⁶.

4.4.6 Picking candidate biomarkers specific to incipient MCI and AD and preclinical AD

Proteomic expression change was seen in a surprisingly large number of plasma proteins in normal ageing over the 6 year period of this study (71 DEPs, Table 4.2). However, even after these were excluded from the MCI and AD longitudinal analyses, I was still left with long lists of proteins (66 and 60 MCI and AD specific DEPs, respectively), and similarly large numbers in the cross-sectional analyses (89 and 70 MCI and AD specific DEPs respectively). Such abundance presents a dilemma as to which DEPs might be ideal biomarker candidates. One approach is to select potential candidates based on consistency of fold-change in both longitudinal and cross-sectional analyses for each incipient AD (Table 4.10) and MCI (Table 4.11) and preclinical AD groups (Figure 4.9D and Table 4.13). These are much shorter lists but are likely much stronger candidates for future validation work. It is of note that most DEPs in the incipient MCI and AD groups, are upregulated (~ 66% each). In contrast, DEPs with consistent fold changes in the preclinical and incipient AD groups are mostly downregulated, with only 2 DEPS (< 20%) upregulated, these two being the cell-matrix protein CH13L1 and the proteasome 20S core protein PSMB2.

4.5 Conclusion

The in-depth plasma proteomics analysis, in combination with longitudinal and cross-sectional analyses of an older age cohorts, ageing with normal cognition or progressing to MCI or AD. These findings revealed changes common to ageing regardless of diagnosis, and molecular similarities and differences between AD and MCI, as well as some putative dementia specific plasma biomarkers for clinical and preclinical AD. The protein changes are consistent across two independent search engines, paving the path for future research on ageing and AD dementia biomarker identification. The considerably larger pool of preclinical AD associated differentially expressed proteins may in part be evidence of pathology in progress, in addition to providing a large pool of putative early biomarkers. Apart from the well know CSF AD biomarker (CHI3L1) we propose a list of novel markers, including PSMB2, PAM, ALDOB, TF, MAPRE2 which may be potential preclinical biomarkers for the identification of AD dementia.

Chapter 5

Impact of *APOE* ϵ 3 and ϵ 4 genotypes on plasma proteome signatures in Alzheimer's disease

Manuscript

5.1 Introduction:

One of the barriers to developing effective therapies for Alzheimer's disease (AD), the most common cause of dementia, lies in the lack of a comprehensive understanding of the brain mechanisms leading to neurodegeneration. One key knowledge gap in understanding how genetic risk factors contribute to disease pathogenesis. There are numerous genetic risk factors for developing sporadic AD, the strongest of which is the apolipoprotein E epsilon 4 allele inheritance (*APOE* ϵ 4)^{13,148-150}. Three alleles, i.e., *APOE* ϵ 2, *APOE* ϵ 3, and *APOE* ϵ 4, result in six possible genotypes (*APOE* 2/2, 2/3, 3/3, 2/4, 3/4, and 4/4). These three polymorphic alleles, i.e., ϵ 2, ϵ 3, and ϵ 4, have a worldwide frequency of 8.4%, 77.9%, and 13.7%, respectively^{13,14}. Recent studies reported that approximately 65% of individuals with late-onset familial and sporadic AD bear the *APOE* ϵ 4 allele⁴. One copy of *APOE* ϵ 4 is associated with a threefold increase in disease risk, while two copies are associated with a more than tenfold increase in risk¹⁵.

Emerging data suggest that *APOE* ϵ 4 is involved in several functions, including metabolism, neuroinflammation, impaired amyloid clearance, transport, synaptogenesis, and glucose, lipid, and cholesterol metabolism in the brain^{327,328 329}. In animal and cellular models, *APOE* ϵ 4 has been linked to decreased cellular plasticity¹⁷⁰. In addition, *APOE* plays a critical role in lipid transport and cholesterol homeostasis in the brain, as it does peripherally^{13,151}. In the CNS, *APOE* is mostly expressed in astrocytes, and it facilitates the transportation of cholesterol to neurons by binding to LDLR family members, known as APOE receptors. *APOE* ϵ 4 has been found to be hypolipidated and less effective at inducing cholesterol efflux than *APOE* ϵ 3, implying that the pathological effects of *APOE* may be associated with lipid metabolism¹⁴⁹. However, approximately 25%–40% of patients with AD do not carry the *APOE* ϵ 4 allele, and the pathophysiological mechanisms underlying AD are less evident in these individuals^{13,152}.

Unbiased proteomics analysis permits the simultaneous evaluation of many molecular processes in patients. To explore this, several research studies have used a CSF proteomics technique and described protein signatures linked with AD across the cognitive range^{330,331}. Proteomics investigations on readily available fluids such as serum/plasma, on the other hand, are underutilized. To gain a better understanding of how *APOE* genotypes may influence AD pathology, I used a plasma proteomic approach to test the hypothesis that protein signatures

can be detected that show *APOE* genotype-dependent associations with AD. The main objective of this study was to understand better the changes in plasma that may contribute to disease pathogenesis in AD and how *APOE* ϵ 3 and *APOE* ϵ 4 contribute to biomarker profiles in AD. This study included AD and cognitively normal age-matched control carriers of the *APOE* ϵ 4 allele and AD and control homozygous *APOE* ϵ 3 carriers. Furthermore, all AD (whether ϵ 3 or ϵ 4) were Pittsburgh compound B (PiB) positron emission tomography (PET) positive (high or very high), whereas all controls (whether ϵ 3 or ϵ 4) were PiB PET negative.

5.2 Materials and Methods:

5.2.1 Cohort and samples

Plasma samples were obtained from the Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) participants aged 70-90 years³³². The University of Melbourne Human Research Ethics Committee approved the collection of the AIBL cohort, while the UNSW Human Research Ethics Committee approved the current study. All work complied with the Declaration of Helsinki guidelines.

In total, I profiled 40 human plasma samples using label-free proteomics in the following four groups: **1.** *APOE* ϵ 4/ ϵ 3 carriers without AD are denoted as CTRLE4, **2.** *APOE* ϵ 4/ ϵ 4 carriers with AD symptoms denoted as ADE4 **3.** *APOE* ϵ 3/ ϵ 3 carriers without AD denoted as CTRLE3 and **4.** *APOE* ϵ 3/ ϵ 3 carriers with AD symptoms are denoted as ADE3.

5.2.2 Depletion of high abundant proteins using Human 14 (Hu14) immunoaffinity Columns:

The protocol for removing plasma high abundance proteins followed by fractionating the low abundance proteins was adapted from a previously published approach³⁸. The top 14 high-abundance plasma proteins (albumin, IgG, antitrypsin, IgA, transferrin, haptoglobin, fibrinogen, α -2-macroglobulin, α -1-acid glycoprotein, IgM, apolipoprotein AI, apolipoprotein AII, complement C3, and transthyretin) were depleted using an Hu14 column (4.6 x 100 mm, Agilent). The plasma (50 μ L) was diluted with 150 μ L of buffer A (1:4 dilutions, as recommended by Agilent Technologies), and then filtered to remove particulates using a 0.45

µm spin filter (Spin-X centrifuge tube filter, 0.45 µm Cellulose Acetate, Merck, Germany). Samples were then injected (100 µL) onto the Hu14 column. Chromatography and fraction collection was performed on an Agilent 1290 UPLC system (Agilent, Santa Clara, CA) with Hu14 buffers A and B purchased from Agilent (Santa Clara, CA), and manufacturer's instructions followed for protein binding and elution (Agilent, Santa Clara, CA). The low abundance protein fraction was further fractionated by 1D SDS PAGE and analyzed using LC-MS/MS. Participant demographics are shown in Table 5.1. Each sample consisted of ten SDS PAGE fractions, so a total of 400 LC-MSMS runs were performed to ensure adequate coverage of the plasma proteome.

5.2.3 Fractionation of low abundance proteins using 1D-SDS PAGE

I followed a previously published procedure for SDS-PAGE, band cutting, trypsin digestion, sample preparation, and mass spectrometric analysis³⁸. Equal amounts of total protein (50 µg) from the HU14 depleted plasma were filtered using Amicon ultra 3kDa centrifugal filter units (MERCK, New Jersey, USA), dried in speed vac (ThermoFisher, Massachusetts, USA), and reconstituted to a final volume of 20 µL by adding 5 µL LDS sample buffer Invitrogen NuPAGE (ThermoFisher, Massachusetts, USA), 2 µL reducing agent Invitrogen NuPAGE (ThermoFisher, Massachusetts, USA), and 13 µL deionized water (MilliQ). After briefly heating samples (10 minutes, 70°C), they were separated by 1D SDS/PAGE using Invitrogen NuPAGE 4-12% Bis-Tris midi gels (ThermoFisher Scientific, Massachusetts, USA) and Invitrogen MES running buffer using the manufacturer's instructions (ThermoFisher Scientific, MA, USA). The gel was then stained using colloidal coomassie G250 (Figure 5.2ii). The protein lanes were cut into gel bits by destaining using a 24-band lane cutting blade. The gel bands were concatenated into ten vials for in-gel trypsin digestion, peptide recovery, and label-free LCMSMS quantification. A total of 10 biological replicates (subjects) were used per group, i.e., *ADE4*, *CTRL4*, *ADE3*, *CTRL3*.

5.2.4 Computational Analysis

Two search engines were used to analyze the raw files, including ProteomeDiscoverer v2.4 (Thermo Fisher Scientific, Waltham, MA) and Scaffold Q+ software v 4.11.0 (Proteome Software, Portland, OR). A minimum of two unique peptides per protein were prerequisites for protein identification and quantitation for both data analysis software. In conjunction with the reversed decoy and frequent contaminant sequences, all search engines used the UniProt Homo

sapiens (human) database for MS and MS/MS spectral mapping. Five parts per million (5ppm) mass tolerance was used to match peaks to theoretical ion series. The false discovery rate (FDR) was set to 1% to ensure that only highly confident protein identifications were made. Trypsin was chosen as the specific enzyme, with a maximum of two missed cleavages. Variable acetylation at the N-terminus of proteins, methionine oxidation, and fixed carbamidomethylation of cysteines was used for all database searches. All parameters were kept consistent between the two search engines. The software used enabled the simultaneous application of two distinct approaches for label-free quantification: peak area integration (PD2.4) and spectral counting (Scaffold).

Each AD group required independent proteomics data processing using the PD2.4 and Scaffold search engines. In Figure 5.2i, scatter plots and regression analysis comparing PD2.4 versus scaffold fold-change are displayed. To identify proteins with significant expression differences between groups, I used the following inclusion criteria: proteins quantified in >5 individuals, proteins identified with a minimum of two peptides per protein, and consistent direction of protein fold change across two bioinformatics platforms using orthogonal quantification approaches (peak area ratio with PD2.4 and spectral counting with Scaffold) with a fold change of at least 20% (≤ 0.08 and ≥ 1.2) in both search engines were selected for the analysis.

5.2.5 Bioinformatics Analysis

Bioinformatics analyses were performed using RStudio version 1.2.5033 and R version 3.6.3 to create heatmaps and volcano plots, using the heatmap function and ggplot2 package. Gene ontology (GO) and enrichment plot analysis were performed using Bioconductor's GOstats and DOSE package. Results from the gene ontology analysis were only studied if more than two genes from the experimental data set were included with a particular term. Volcano plot analysis was performed using the Enhanced Volcano package from Bioconductor³³³. Venn diagrams were plotted using Venny 2.114. I used differentially expressed proteins (DEPs) to compare biological processes and pathways affected in AD versus control using GO enrichment analysis on the STRING (v11.0) explored gene interaction and co-expression patterns for differentially expressed genes (DEGs).

5.3 Results:

5.3.1 Sample characteristics by *APOE* ϵ 3 and *APOE* ϵ 4 genotype

I profiled 40 human plasma samples using label-free proteomics in the following four groups: (1) *APOE* ϵ 4/ ϵ 3 carriers, cognitively normal controls with negative PiB PET denoted as CTRL E4 (2) *APOE* ϵ 4/ ϵ 4 carriers with AD symptoms and positive PiB PET denoted as ADE4 (3) *APOE* ϵ 3/ ϵ 3 carriers, cognitively normal controls with negative PiB PET denoted as CTRL E3 (4) *APOE* ϵ 3/ ϵ 3 carriers with AD symptoms and positive PiB PET denoted as ADE3. I identified 1,055 proteins (false discovery rate <1%) with 23,242 total peptides using Proteome Discoverer 2.4 (PD2.4) search engine and 800 proteins using Scaffold. More than 700 identified proteins were common in both software techniques.

An overview of the study populations and proteomic workflow is shown in Figure 5.1A. Box plots show the similar distribution and protein abundance variation across all 40 plasma samples (Figure 5.1B). The overall similarity of low abundance proteins across samples is also evident by SDS PAGE (Figure 5.2ii). Unsupervised hierarchical clustering analysis (HCA) of grouped abundances (data from PD2.4 software processing) is presented in Figure 5.1C. It shows that control and AD samples of both *APOE* ϵ 3 and *APOE* ϵ 4 carriers cluster together more closely based on diagnosis sample type (i.e., control vs. AD) rather than *APOE* allele type. Nevertheless, distinct proteomic profiles are observed in each of the four groups since all heat maps are quite distinct (Figure 5.1C). The scatter plot depicting AD and CTRL data points analyzed on both PD2.4 and Scaffold is shown in Figure 5.1D. The detailed scatter plots and density plots were plotted using the complete list of proteins from both the search engines in Figure 5.2i. In Figure 5.2ii, the scatter plots show only the differentially expressed proteins (DEPs) using all the analyses, i.e., **A.** ADE3/CTRL E3, **B.** ADE4/CTRL E3, **C.** ADE4/ADE3, **D.** CTRL E4/CTRL E3, and **E.** ADE4/CTRL E4. The DEPs with a similar direction of change using both orthogonal quantification techniques, PD2.4 peak area ratio, and scaffold spectral counting are shown in Figure 5.1E. The bar graph shows the total number of proteins upregulated and downregulated in five comparisons, i.e., ADE3/CTRL E3, ADE4/CTRL E3, CTRL E4/CTRL E3, ADE4/CTRL E4, and ADE4/ADE3 (Figure 5.1E).

Table 5.1: Demographics of the participants.

Total participants	Control <i>APOEε3</i>	Control <i>APOEε4</i>	AD <i>APOEε3</i>	AD <i>APOEε4</i>	Kruskal- Wallis statistic	Kruskal-Wallis P value
Total participants in each wave	10	10	10	10	NA	NA
Age in years mean±SD (CV%)	72.40±5.0 (6.9%)	72.10±6.2 (8.6%)	70.50±5.6 (8.03%)	70.70±6.5 (9.28%)	0.42	0.93
Education (years)	12.90±3.2 (24.91%)	14.80±2.3 (15.86%)	11.60±3.1 (27.02%)	12.70±2.2 (17.43%)	7.47	0.05
Sex (n)	F=5, M=5	F=6, M=4	F=6, M=4	F=5, M=5	NA	NA
<i>APOE</i> status	E3/3	E3/4	E3/3	E3/4	NA	NA
MMSE mean±SD (CV%)	29.00±1.0 (3.63%)	29.50±0.5 (1.78%)	19.50±4.9 (25.5%)	21.40±7.6 (35.8%)	30.51	0.00
Hypertension in number of participants (%)	50%	40%	60%	30%	NA	NA
Diabetes in number of participants (%)	10%	10%	0%	10%	NA	NA
Cholesterol (mmol/L) mean±SD (CV%)	5.29±1.2 (22.81%)	5.45±1.5 (27.83%)	5.65±1.0 (18.34%)	5.63±1.1 (20.83%)	0.53	0.91
Triglyceride (mmol/L) mean±SD (CV%)	1.02±0.2 (26.06%)	1.43±0.52 (36.72%)	1.23±0.50 (40.50%)	1.63±1.45 (89.25%)	3.96	0.26
HDL-Chol (mmol/L) mean±SD (CV%)	1.72±0.53 (31.03%)	1.52±0.48 (32.07%)	1.67±0.42 (25.10%)	1.47±0.37 (25.69%)	0.61	1.81
LDL-Chol (mmol/L) mean±SD (CV%)	3.09±0.93 (30.38%)	3.27±1.21 (37.06%)	3.41±0.94 (27.64%)	3.40±1.20 (35.48%)	0.65	0.88
Urea (mmol/L) mean±SD (CV%)	6.40±1.17 (18.43%)	5.98±8.90 (21.89%)	5.36±0.92 (17.18%)	6.27±2.09 (33.14%)	3.69	0.29

SD= standard deviation; cv= coefficient of variations

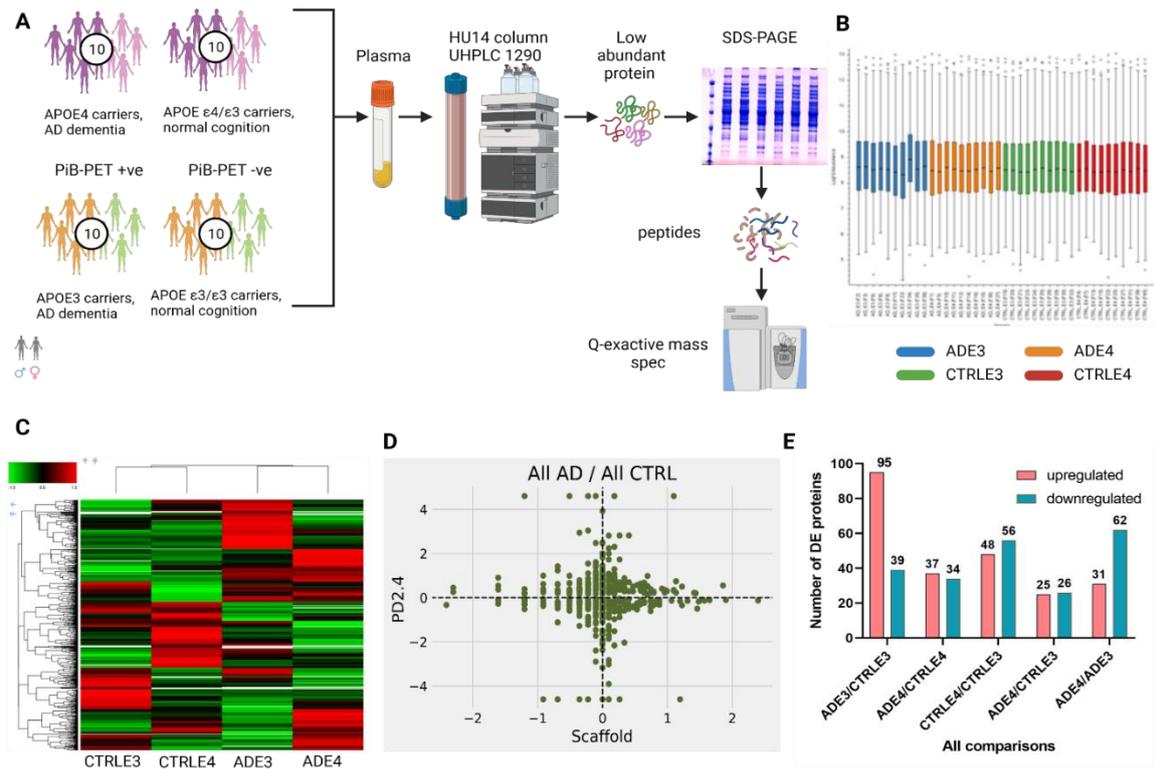
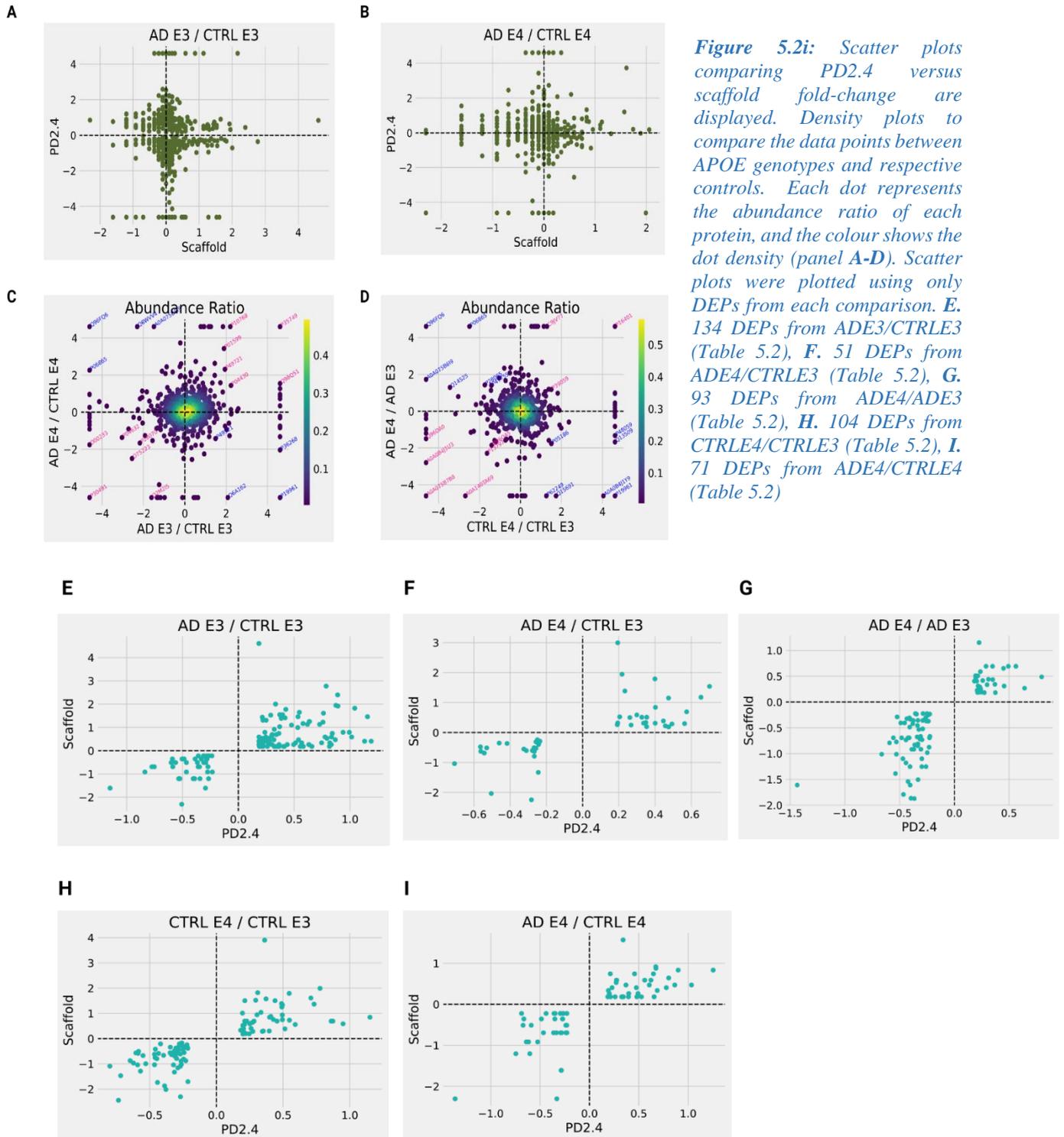


Figure 5.1: Workflow of plasma proteome profiling and comparison of *APOE* ϵ 3 and *APOE* ϵ 4 genotype. **A.** Overview of the study populations and schematic proteomic workflow. Dark and light shades represent male and female subjects, respectively. The flow diagram outlines steps of sample preparation through to data acquisition. **B.** Box-and-whisker plots of abundance values of all 40 individual samples. The small horizontal line within each box denotes the median value, and the upper and lower ranges (whiskers) indicate the 5 and 95 percentiles of the abundance values, respectively (output from ProteomeDiscoverer 2.4 software). **C.** Hierarchical clustering analysis (HCA) of the whole proteome of 40 individuals (10 individuals in each category, i.e., ADE3, control E3, ADE4, and control E4) (output from ProteomeDiscoverer 2.4 software). **D.** Scatter plot depicting the positioning of all AD and CTRL data points analysed on both PD2.4 and Scaffold. **E.** Global analyses of proteomic changes in longitudinal groups. Bar graph showing the total number of proteins upregulated (pink) and downregulated (blue) in ADE3/CTRL E3, ADE4/CTRL E4, CTRL E4/CTRL E3, ADE4/CTRL E3, ADE4/ADE3 and comparison groups. The numbers at the top of each bar indicate the number of differentially expressed proteins (DEPs) in that category. This data was based on the criteria for DEP selection outlined in the method (i.e., $\geq 20\%$ fold-change, the consistent direction of fold change in two orthogonal quantification methods, change identified in >5 individuals per group).



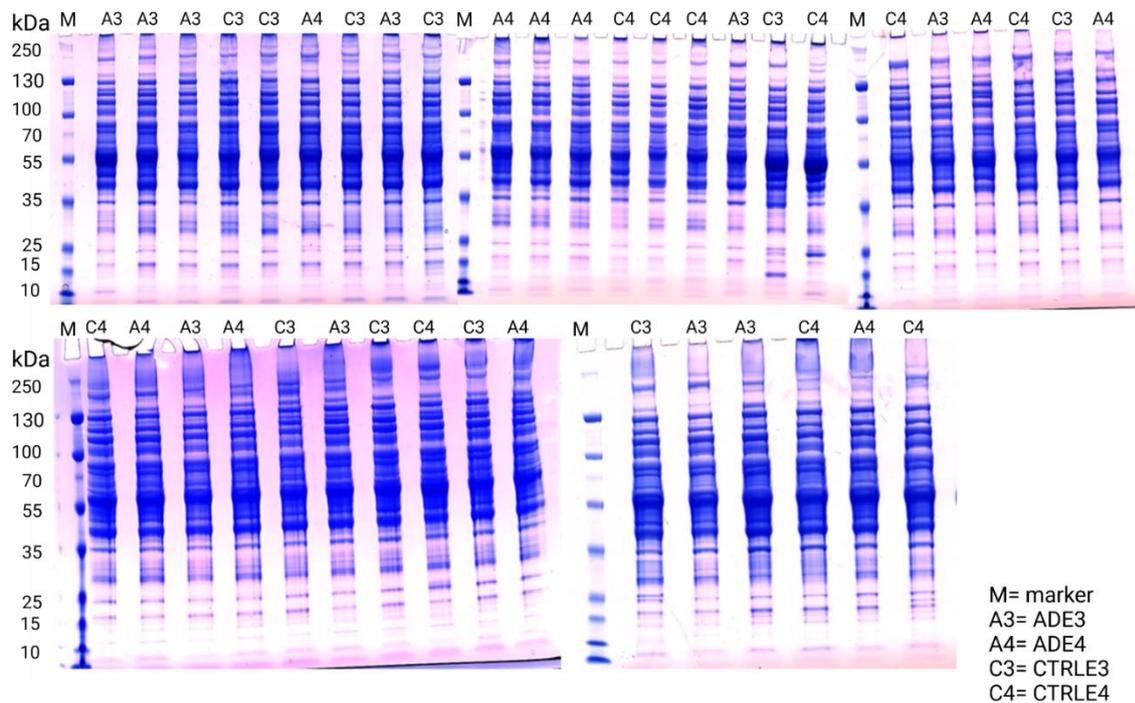


Figure 5.2ii: Images of NuPAGE LDS gels of depleted plasma containing low abundant plasma proteins (LAP), the flow-through from the HU14 column. Each gel lane contained equal protein loading (50 ug total proteins were loaded per gel lane). I have used 10 LAP samples from each sample A3 denotes ADE3, A4 denotes ADE4, C3 denotes control E3, and C4 denotes control E4.

5.3.2 Overall plasma proteome changes in AD vs controls in *APOEε3* carriers:

A heatmap of the total of 134 proteins that were differentially expressed (95 upregulated and 39 downregulated) in ADE3 relative to E3 controls (ADE3/CTRL E3) is shown in two panels (Figure 5.3Ai and 5.3Aii) for better visibility of the protein acronyms and fold changes. The DEPs with the highest fold change in the ADE3/CTRL E3 group are shown in a volcano plot, using the PD2.4 abundance ratios (Figure 5.3B), with the complete list of DEPs shown in Table 5.2 and Table 5.3. A subset of 65 DEPs (48 upregulated and 17 downregulated) was unique to the ADE3/CTRL E3 group and did not exhibit differential expression in other comparison groups, including ADE4/ADE3, ADE4/CTRL E4, or CTRL E4/CTRL E3 comparisons (Figure 5.4 and Table 5.2).

These unique DEPs were further manually categorized, based on gene ontology, to various biological activities, using data from the PD2.4 analyses (Figure 5.4E-H). A more detailed GO enrichment analysis using STRING software was also performed. The three groups with the most significant proportion of DEPs included metabolism (38%), protein binding (11%), and

formation of the cornified envelope (9%) (Figure 5.4F). Interestingly, from this list, two DEPs, serum amyloid P-component (APCs) and lactotransferrin (LTF), have been linked to the formation of amyloid fibrils³³⁴. Two other DEPs were linked to Alzheimer's disease pathway^{335,336}: peroxiredoxin-2 (PRDX2) and extracellular superoxide dismutase (SOD3). Other pathways implicated in neurodegenerative disease included; MAPK activation (6 DEPs; Actin cytoplasmic 1 (ACTB), Annexin A1 (ANXA1), voltage-dependent calcium channel subunit alpha-2/delta-1 (CACNA2D1), and 14-3-3 protein beta/alpha (YWHAB), Fibronectin (FN1), and Proteasome subunit alpha type-4 (PSMA4) and HIF1 signalling pathways (5 DEPs; glyceraldehyde-3-phosphate dehydrogenase (GAPDH), insulin-like growth factor-binding protein 2 (IGFBP2), phosphoglycerate kinase 1 (PGK1), transferrin receptor protein 1 (TFRC), and metalloproteinase inhibitor 1 (TIMP1), (Figure 5.4F and Table 5.3). Previous studies have demonstrated the role of HIF1 signalling in neurodegenerative disease^{337,338}. Given that *APOE* ϵ 3 is the most common population variant, these DEPs may provide insights into the underlying processes related to AD, but not due to the presence of the *APOE* ϵ 4 variant.

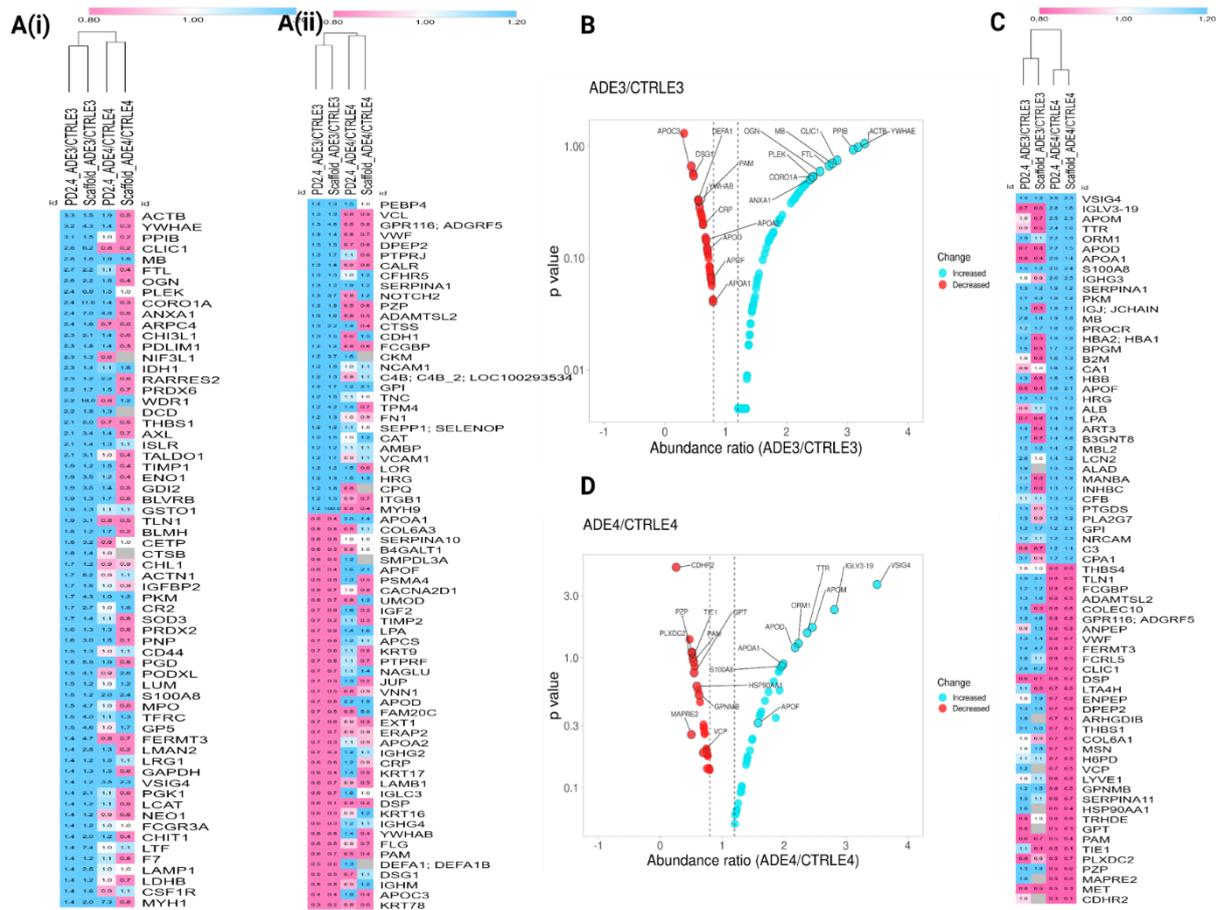


Figure 5.3: Global analyses of proteomic changes in ADE3/CTRLE3 and ADE4/CTRLE4 analysis.

Ai and *ii*. Heatmap showing a total of 134 proteins differentially expressed, including 95 upregulated and 39 downregulated DEPs in ADE3 relative to E3 controls based on the ADE3/CTRLE3. The heatmap is given in two panels (with *Ai* and *Aii* continuing one after the other) to make protein acronyms and fold changes legible. Expression changes of the same proteins in the ADE4/CTRLE4 group are shown alongside for comparison. *B*. Volcano plot of DEPs in the ADE3/CTRLE3 group, using the abundance ratios from PD2.4, which had at least a 20% fold change, the consistent direction of fold-change across the two software platforms (Scaffold and PD2.4), and were identified in 50% or more of subjects. To avoid crowding, I have highlighted only a few DEPs, with a complete list of DEPs shown in the heatmap and shown with greater detail in Table 5.2 and 5.3. *C*. Heatmap showing a total of 71 proteins differentially expressed, including 37 upregulated and 34 downregulated, in ADE4 relative to E4 controls based on the ADE4/CTRLE4. Expression change of the same proteins in the ADE3/CTRLE3 group are shown alongside for comparison. *D*. Volcano plot of DEPs in ADE4/CTRLE4 was created using the abundance ratios from PD2.4, which had at least a 20% fold change, the consistent direction of fold-change across the two software platforms (Scaffold and PD2.4), and were identified in 50% or more of subjects. To avoid crowding, I have highlighted only a few DEPs, a complete list of DEPs shown in the heatmap, and Table 5.2. *E*. Heatmap showing a total of 51 proteins differentially expressed, including 25 upregulated and 26 downregulated DEPs in ADE4 relative to E3 controls based on the ADE4/CTRLE3 data shown in Table 5.2. *F*. Volcano plot of DEPs in ADE4/CTRLE3, using the abundance ratios from PD2.4, which had at least a 20% fold change, the consistent fold-change direction across the two software platforms (Scaffold and PD2.4), and were identified in 50% or more of subjects. To avoid crowding, I have highlighted only a few DEPs, with a complete list of DEPs shown in the heatmap and more detail in Table 5.2. *G*. Heatmap showing a total of 104 proteins differentially expressed, including 48 upregulated and 56 downregulated in control E4 relative to E3 controls based on the CTRLE4/CTRLE3 data shown in Table 5.2. The heatmap is split into two panels (*Gi* and *Gii*) to make protein acronyms and fold changes legible. The *H*. Volcano plot of DEPs in CTRLE4/CTRLE3 using the abundance ratios from PD2.4, which had at least a 20% fold change, the consistent fold-change direction across the two software platforms (Scaffold and PD2.4), and were identified in 50% or more of subjects. To avoid crowding, I have highlighted only a few DEPs, with a complete list of DEPs shown in the heatmap and more detail in Table 5.2.

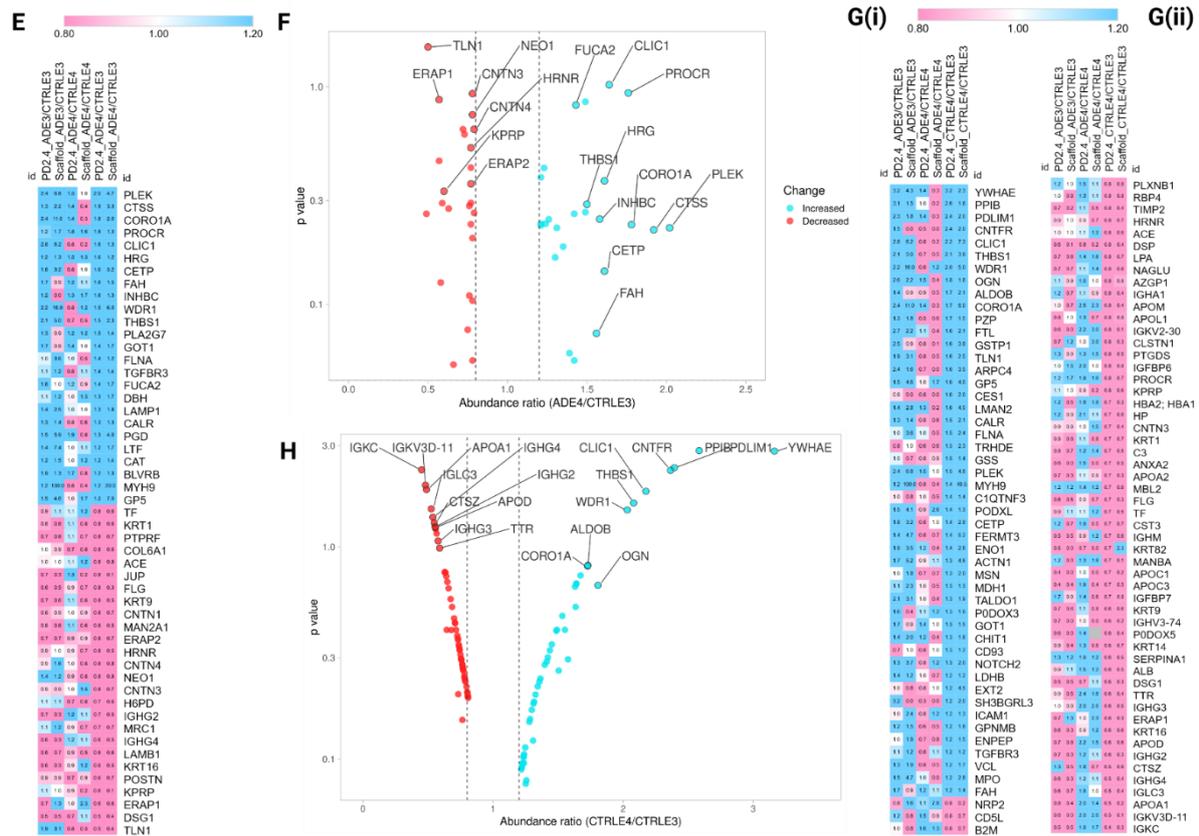


Table 5.2. The final list of differentially expressed proteins (DEPs) in all the comparisons analysed. This list contains DEPs those quantified in >5 individuals, proteins identified with a minimum of two peptides/protein, the consistent direction of protein fold change across two bioinformatics platforms with orthogonal quantification approaches (peak area ratio with PD2.4 and spectral counting with Scaffold) with a fold change of at least 20% (≤ 0.8 and ≥ 1.2) in both search engines.

Comparisons analysed	Protein Gene Symbol
A list of 71 DEPs were identified in ADE4 relative to control E4	Total protein number = 71 (37 Upregulated, 34 Downregulated) VSIG4, IGLV3-19, APOM, TTR, ORM1, APOD, APOA1, S100A8, IGHG3, SERPINA1, PKM, IGJ;JCHAIN, MB, PROCR, HBA2;HBA1, BPGM, B2M, CA1, HBB, APOF, HRG, ALB, LPA, ART3, B3GNT8, MBL2, LCN2, ALAD, MANBA, INHBC, CFB, PTGDS, PLA2G7, GPI, NRCAM, C3, CPA1, THBS4, TLN1, FCGBP, ADAMTSL2, COLEC10, GPR116;ADGRF5, ANPEP, VWF, FERMT3, FCRL5, CLIC1, DSP, LTA4H, ENPEP, DPEP2, ARHGDI, THBS1, COL6A1, MSN, H6PD, VCP, LYVE1, GPNMB, SERPINA11, HSP90AA1, TRHDE, GPT, PAM, TIE1, PLXDC2, PZP, MAPRE2, MET, CDHR2

<p>A list of 105 DEPs, identified in control E4 relative to control E3</p>	<p>Total protein number = 105 (48 Upregulated, 57 Downregulated) YWHAE, PPIB, PDLIM1, CNTFR, CLIC1, THBS1, WDR1, OGN, ALDOB, CORO1A, PZP, FTL, GSTP1, TLN1, ARPC4, GP5, CES1, LMAN2, CALR, FLNA, TRHDE, GSS, PLEK, MYH9, C1QTNF3, PODXL, CETP, FERMT3, ENO1, ACTN1, MSN, MDH1, TALDO1, GOT1, CHIT1, CD93, NOTCH2, LDHB, EXT2, SH3BGRL3, ICAM1, GPNMB, ENPEP, TGFBR3, VCL, MPO, FAH, NRP2, CD5L, B2M, PLXNB1, RBP4, TIMP2, HRNR, ACE, DSP, LPA, NAGLU, AZGP1, IGHA1, APOM, APOL1, IGKV2-30, CLSTN1, PTGDS, IGFBP6, PROCR, KPRP, HBA2;HBA1, HP, CNTN3, KRT1, C3, ANXA2, APOA2, MBL2, FLG, TF, CST3, IGHM, KRT82, MANBA, APOC1, APOC3, IGFBP7, KRT9, IGHV2-26, , KRT14, SERPINA1, ALB, DSG1, TTR, IGHG3, ERAP1, KRT16, APOD, IGHG2, CTSZ, IGHG4, IGLC3, APOA1, IGKV3D-11, IGKC</p>
<p>A list of total 134 DEPs were identified in ADE3 relative to control E3</p>	<p>Total protein number = 89 (95 Upregulated, 39 Downregulated) ACTB, YWHAE, PPIB, CLIC1, MB, FTL, OGN, PLEK, CORO1A, ANXA1, ARPC4, CHI3L1, PDLIM1, NIF3L1, IDH1, RARRES2, PRDX6, WDR1, DCD, THBS1, AXL, ISLR, TALDO1, TIMP1, ENO1, GDI2, BLVRB, GSTO1, TLN1, BLMH, CETP, CTSB, CHL1, ACTN1, IGFBP2, PKM, CR2, SOD3, PRDX2, PNP, CD44, PGD, PODXL, LUM, S100A8, MPO, TFRC, GP5, FERMT3, LMAN2, LRG1, GAPDH, VSIG4, PGK1, LCAT, NEO1, FCGR3A, CHIT1, LTF, F7, LAMP1, LDHB, CSF1R, MYH1, PEBP4, VCL, GPR116;ADGRF5, VWF, DPEP2, PTPRJ, CALR, CFHR5, SERPINA1, NOTCH2, PZP, ADAMTSL2, CTSS, CDH1, FCGBP, CKM, NCAM1, C4B;C4B_2;LOC100293534, GPI, TNC, TPM4, FN1, SEPP1;SELENOP, CAT, AMBP, VCAM1, LOR, HRG, CPQ, ITGB1, MYH9, APOA1, COL6A3, SERPINA10, B4GALT1, SMPDL3A, APOF, PSMA4, CACNA2D1, UMOD, IGF2, TIMP2, LPA, APCS, KRT9, PTPRF, NAGLU, JUP, VNN1, APOD, FAM20C, EXT1, ERAP2, APOA2, IGHG2, CRP, KRT17, LAMB1, IGLC3, DSP, KRT16, IGHG4, YWHAB, FLG, PAM, DEFA1;DEFA1B, DSG1, IGHM, APOC3, KRT78</p>
<p>A list of 51 DEPs, identified in AD E4 relative to control E3</p>	<p>Total protein number = 51 (25 Upregulated, 26 Downregulated) PLEK, CTSS, CORO1A, PROCR, CLIC1, HRG, CETP, INHBC, FAH, THBS1, PLA2G7, WDR1, FUCA2, FLNA, TGFBR3, GOT1, LAMP1, CALR, DBH, PGD, GP5, LTF, CAT, MYH9, BLVRB, CNTN4, CNTN1, PTPRF, FLG, CNTN3, NEO1, ACE, KRT1, ERAP2, KRT9, HRNR, TF, COL6A1, MAN2A1, JUP, IGHG2, H6PD, MRC1, POSTN, KPRP, LAMB1, KRT16, IGHG4, ERAP1, TLN1, DSG1</p>
<p>A list of 93 DEPs, identified in ADE4 relative to ADE3</p>	<p>Total protein number = 93 (31 Upregulated, 62 Downregulated) APOM, CRP, ELTD1;ADGRL4, SELENBP1, TTR, LPA, PON1, PON3, CD93, ART3, B3GNT8, C1QTNF3, B4GALT1, BPGM, DBH, FCN3, ALDOC, CALR, B2M, VNN1, IGHG3, HBA2;HBA1, ADAMTSL4, NAGLU, PLA2G7, C1QB, CNDP1, HEG1, PKHD1L1, APOF, LTA4H, ERAP1, ALDOA, FCRL5, TIMP1, SOD3, CD109, CHI3L1, ALAD, IGFBP7, BLMH, VCAM1,</p>

	IGFBP2, CPA1, COL6A1, MRC1, ISLR, P4HB, L1CAM, B3GNT1;B4GAT1, VCL, TLN1, THBS1, CKM, SPP2, FCGBP, ADAMTSL2, CD163, H6PD, CD44, CFHR2, GPNMB, ENO1, ARHGDI1, ICAM2, DPEP2, VWF, CAPN1, KPRP, A2M, CAP1, AXL, PGD, SERPINA11, MET, LILRA2, SMPDL3A, PZP, WDR1, TALDO1, NEO1, GANAB, MAN2A2, GSTP1, MB, WARS, CR2, TUBB1, CLIC1, ILK, PKM, ANXA1, STATH
--	--

Table 5.3: This table shows the unique 65 DEPs identified in ADE3 relative to control E3. These DEPs did not clear the above criteria in ADE4/CTRLE4, CTRLE4/CTRLE3 and ADE4/CTRLE3 groups, so they are considered the DEPs unique the ADE3/CTRLE3 group.

Accession	Description	Gene Symbol	PD2.4_ADE3/CTRLE3	Scaffold_ADE3/CTRLE3	PD2.4_Adj. P-Value:ADE3/CTRLE3	Scaffold_T-Test (p-value): ADE3/CTRLE3
P60709	Actin, cytoplasmic 1	ACTB	3.285	1.5	0.08755265	0.43
P04083	Annexin A1	ANXA1	2.394	7	0.311744894	0.24
P36222	Chitinase-3-like protein 1	CHI3L1	2.318	2.1	0.346965541	0.33
Q9GZT8	NIF3-like protein 1	NIF3L1	2.297	1.3	0.354832127	0.87
O75874	Isocitrate dehydrogenase [NADP] cytoplasmic	IDH1	2.267	1.4	0.370922353	0.73
Q99969	Retinoic acid receptor responder protein 2	RARRES2	2.26	1.2	0.374773731	0.78
P30041	Peroxiredoxin-6	PRDX6	2.211	1.7	0.406547747	0.42
P81605	Dermcidin	DCD	2.173	1.8	0.428967303	0.7
P30530	Tyrosine-protein kinase receptor UF	AXL	2.133	3.4	0.45629486	0.26
O14498	Immunoglobulin superfamily containing leucine-rich repeat protein	ISLR	2.107	1.4	0.467581024	0.47
P01033	Metalloproteinase inhibitor 1	TIMP1	1.931	1.2	0.573462796	0.76
P50395	Rab GDP dissociation inhibitor beta	GDI2	1.887	3.5	0.599785554	0.11
P78417	Glutathione S-transferase omega-1	GSTO1	1.878	1.3	0.606067261	0.69
Q13867	Bleomycin hydrolase	BLMH	1.788	1.2	0.674221443	0.85

P07858	Cathepsin B	CTSB	1.762	1.4	0.686666057	0.82
O00533	Neural cell adhesion molecule L1-like protein	CHL1	1.741	1.2	0.701619316	0.38
P18065	Insulin-like growth factor-binding protein 2	IGFBP2	1.713	1.6	0.714769493	0.1
P20023	Complement receptor type 2	CR2	1.702	2.7	0.723907693	0.24
P08294	Extracellular superoxide dismutase [Cu-Zn]	SOD3	1.669	1.4	0.75310847	0.33
P32119	Peroxiredoxin-2	PRDX2	1.64	1.3	0.776985738	0.33
P00491	Purine nucleoside phosphorylase	PNP	1.609	3	0.800606466	0.2
P16070	CD44 antigen	CD44	1.549	1.3	0.846542904	0.33
P51884	Lumican	LUM	1.509	1.2	0.873405528	0.21
P02786	Transferrin receptor protein 1	TFRC	1.468	4	0.90682001	0.25
P02750	Leucine-rich alpha-2-glycoprotein	LRG1	1.44	1.2	0.919863933	0.33
P04406	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	1.44	1.3	0.919863933	0.68
P00558	Phosphoglycerate kinase 1	PGK1	1.437	2.1	0.922586203	0.23
P04180	Phosphatidylcholine-sterol acyltransferase	LCAT	1.436	1.2	0.924440947	0.55
P08637	Low affinity immunoglobulin gamma Fc region receptor III-A	FCGR3A	1.411	1.2	0.942130326	0.52
P08709	Coagulation factor VII	F7	1.393	1.2	0.953814063	0.13
P07333	Macrophage colony-stimulating factor 1 receptor	CSF1R	1.375	1.6	0.962231806	0.0085
P12882	Myosin-1	MYH1	1.353	2	0.979750564	0.31
Q96S96	Phosphatidylethanolamine-binding protein 4	PEBP4	1.352	1.3	0.979750564	0.62

Q12913	Receptor-type tyrosine-protein phosphatase eta	PTPRJ	1.317	1.7	0.989674877	0.072
Q9BXR6	Complement factor H-related protein 5	CFHR5	1.307	1.3	0.989674877	0.43
P12830	Cadherin-1	CDH1	1.264	1.5	0.989674877	0.44
P06732	Creatine kinase M-type	CKM	1.246	3.7	0.989674877	0.49
P13591	Neural cell adhesion molecule 1	NCAM1	1.246	1.5	0.989674877	0.087
P0COL5	Complement C4-B	C4B; C4B_2; LOC100293534	1.244	1.3	0.989674877	0.034
P24821	Tenascin	TNC	1.24	1.5	0.989674877	0.2
P67936	Tropomyosin alpha-4 chain	TPM4	1.239	4.2	0.989674877	0.18
P02751	Fibronectin	FN1	1.229	1.3	0.989674877	0.095
P49908	Selenoprotein P	SEPP1; SELENO P	1.229	1.2	0.989674877	0.21
P02760	Protein AMBP	AMBP	1.214	1.2	0.989674877	0.0057
P19320	Vascular cell adhesion protein 1	VCAM1	1.213	1.7	0.989674877	0.031
P23490	Loricrin	LOR	1.207	1.2	0.989674877	0.9
Q9Y646	Carboxypeptidase Q	CPQ	1.203	1.8	0.989674877	0.34
P05556	Integrin beta-1	ITGB1	1.202	1.5	0.989674877	0.4
P12111	Collagen alpha-3 (VI) chain	COL6A3	0.789	0.8	0.90682001	0.36
Q9UK55	Protein Z-dependent protease inhibitor	SERPINA10	0.769	0.8	0.876762236	0.27
P15291	Beta-1,4-galactosyltransferase 1	B4GALT1	0.766	0.3	0.868634975	0.38
Q92484	Acid sphingomyelinase-like phosphodiesterase 3a	SMPDL3A	0.762	0.5	0.862319392	0.64
P25789	Proteasome subunit alpha type-4	PSMA4	0.755	0.8	0.850729325	0.86

P54289	Voltage-dependent calcium channel subunit alpha-2/delta-1	CACNA2D1	0.755	0.7	0.850729325	0.36
P07911	Uromodulin	UMOD	0.751	0.7	0.850729325	0.8
P01344	Insulin-like growth factor II	IGF2	0.747	0.6	0.845985558	0.59
P02743	Serum amyloid P-component	APCS	0.735	0.8	0.830352661	0.046
O95497	Pantetheinase	VNN1	0.706	0.5	0.776100017	0.2
Q8IXL6	Extracellular serine/threonine kinase FAM20C	FAM20C	0.698	0.5	0.7601127	0.68
Q16394	Exostosin-1	EXT1	0.693	0.6	0.750519635	0.49
P02741	C-reactive protein	CRP	0.628	0.6	0.62990742	0.25
Q04695	Keratin, type I cytoskeletal 17	KRT17	0.628	0.4	0.62990742	0.61
P31946	14-3-3 protein beta/alpha	YWHAB	0.57	0.6	0.519452427	0.63
P59665	Neutrophil defensin 1	DEFA1; DEFA1B	0.548	0.6	0.466403168	0.69
Q8N1N4	Keratin, type II cytoskeletal 78	KRT78	0.317	0.2	0.049320171	0.41

5.3.3 Proteome changes seen in *APOEε4* carriers:

The 71 DEPs observed in *APOEε4* carriers (ADE4/CTRLE4) included 37 upregulated and 34 downregulated DEPs in ADE4 relative to E4 controls (Table 5.2, Figure 5.1E, and 5.3C). The DEPs in ADE4/CTRLE4 are shown in volcano plot format (Figure 5.3D). A complete list of DEPs in *APOEε4* carriers (ADE4/CTRLE4) is presented in Table 5.2. Further, GO analysis using STRING was performed to obtain the detailed GO enrichment shown in Figure 5.6.

When comparing ADE4 to control E3, a total of 51 DEPs were identified, including 25 upregulated and 26 downregulated proteins (Figure 5.1E, Table 5.2), which are also shown in the heatmap (Figure 5.3E) and a volcano plot (Figure 5.3F). A complete list of DEPs for the ADE4/CTRLE3 comparison is presented in Table 5.2. GO analysis revealed that metabolism (19%), signalling (15%), phagosome (12%), platelet degranulation (12%), and platelet degranulation (10%) were the functional groups containing most ADE4/CTRLE3 DEPs

(Figure 3H). Other GO enrichments include amyloid fibril formation (LTF), apoptosis (DSG1), LDL remodelling (CETP), brain-derived neurotrophic factor signalling pathway (PTPRF) (Figure 5.4H).

While there were some shared DEPs with ADE3/CTRLE3 and ADE4/CTRLE4, I found 29 DEPs (13 upregulated and 16 downregulated) that were uniquely dysregulated in ADE4/CTRLE4 (Figure 5.4D and Table 5.2 and Table 5.4). These AD-related DEPs specific to E4 carriers were manually categorized based on gene ontology, with the majority involved in metabolism (34%), protein binding (21%), and signalling (18%). Other enriched categories in ADE4/CTRLE4 included complement cascade, PI3K-Akt pathway, post-translational modifications (PTMs), protein digestion, and protease inhibitors (Figure 5.4G). The PI3K-Akt pathway, which includes heat shock protein HSP 90-alpha (HSP90AA1), collagen alpha-1(VI) chain (COL6A1), and thrombospondin-4 (THBS4), was one of the distinct pathways dysregulated. Other DEPs included rho GDP-dissociation inhibitor 2 (ARHGDIB) involved in rho GTPase signalling, tyrosine-protein kinase receptor Tie-1 (TIE1) in Rac1/Pak1/p38/MMP-2 pathway, neutrophil gelatinase-associated lipocalin (LCN2) in interleukin-4 and 13 signalling, and lymphatic vessel endothelial hyaluronic acid receptor 1 (LYVE1) in hyaluronan uptake and degradation (Figure 5.4G and Table 5.4).

To explore the role of the *APOE* allele in AD, both ADE4 and ADE3 groups were expressed relative to CTRLE3 (Figure 5.5D). Most of the DEPs in these two comparisons were unique to each group (18 in ADE4/CTRLE3 and 87 in ADE3/CTRLE3; Figure 5.5A), while 28 DEPs were shared, of which the majority (26 DEPs) varied in a similar direction of fold change (Figure 5.5D, Table 5.6). The majority of these DEPs were involved in metabolism (19%), signalling (15%), platelet degranulation (12%), phagosome (12%), and PTM (10%) (Figure 5.4H).

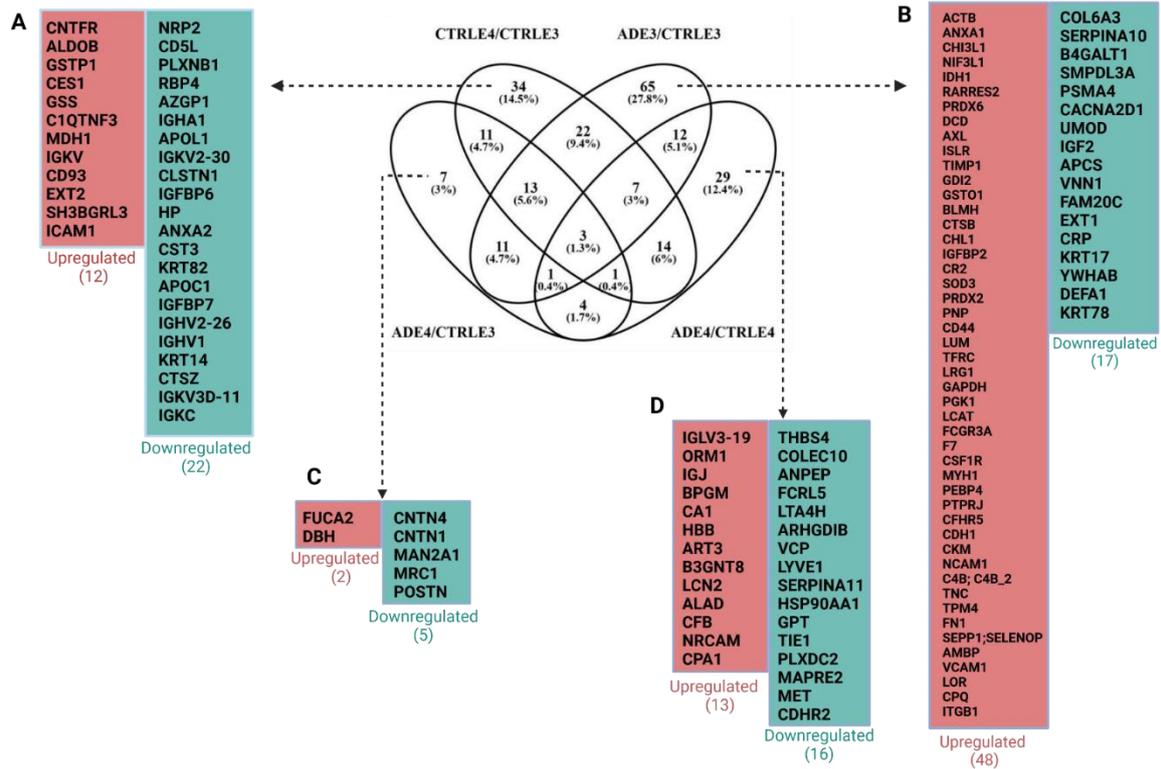


Figure 5.4: Venn diagram of overlapping and unique DEPs in four groups; CTRLE4/CTRLE3, ADE3/CTRLE3, ADE4/CTRLE3, ADE4/CTRLE4, with DEPs unique to each group displayed on the periphery, as follows; A. 34 DEPs unique to the CTRLE4/CTRLE3 group, comprising 12 upregulated and 22 downregulated, in control E4 relative to control E3 (greater detail of the complete list of DEPs identified in the CTRLE4/CTRLE3 group is shown in Table 5.8). This list contains proteins associated with protection against cognitive decline and neuropathology in *APOEε4* carriers who remain cognitively normal. B. 65 DEPs (48 upregulated and 17 downregulated) were unique to ADE3/CTRLE3 (complete DEP list and more detail can be found in Table 5.3). These are DEPs observed in AD subjects who do not carry an *APOEε4* allele, so protein expression changes are associated with AD but unrelated to the E4 allele. C. 7 DEPs (2 upregulated and 5 downregulated), unique to the ADE4/CTRLE3 group (complete DEP list and more detail can be found in Table 5.5). These DEPs may reflect the contribution of the E4 allele to AD since the ADE3/CTRLE3 group does not share them. D. 29 DEPs (13 were upregulated and 16 downregulated) that were explicitly dysregulated in the ADE4/CTRLE4 group this list represents an experimental correction for the presence of the E4 allele by using normal controls which are heterozygous carriers of the E4 allele (complete DEP list and more detail can be found in Table 5.4). Pie charts categorizing all the unique DEPs in each group into their biological processes and molecular pathways based on gene ontology (GO) E. ADE3/CTRLE3, F. CTRLE4/CTRLE3, G. ADE4/CTRLE4, H. ADE4/CTRLE3.

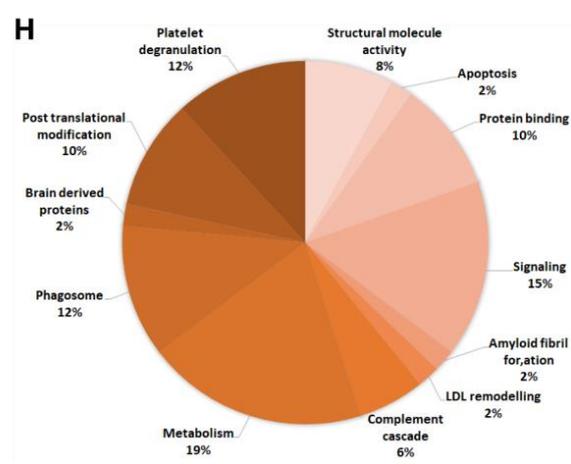
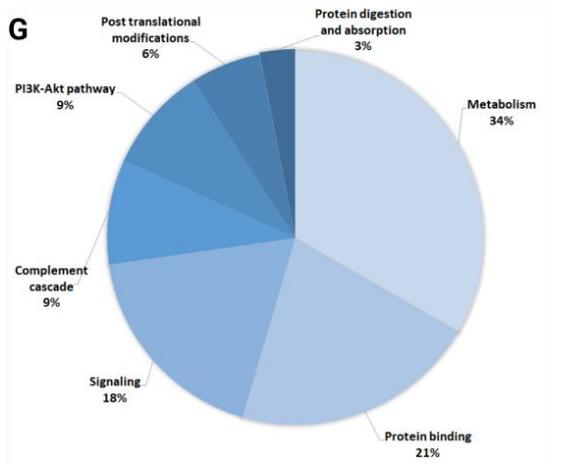
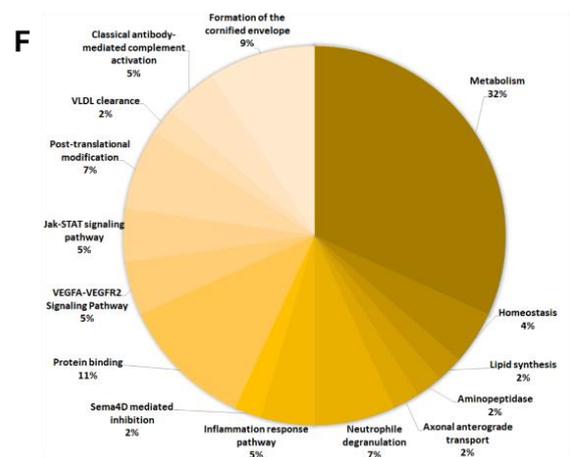
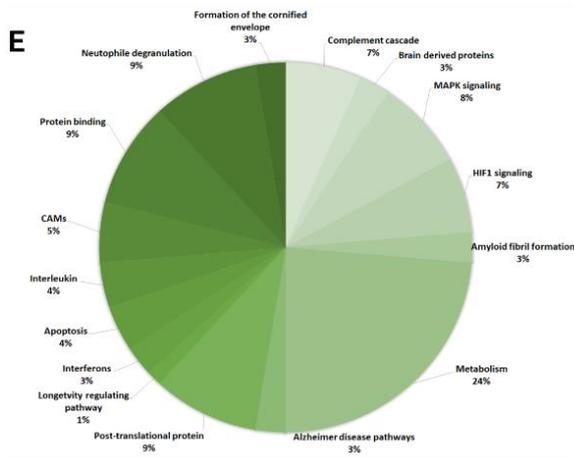


Table 5.4. This table contains the list of 29 unique DEPs identified in ADE4 relative to control E4. These DEPs did not clear the above criteria in ADE4/CTRLE3, ADE3/CTRLE3 and CTRLE4/CTRLE3 groups, so they are considered the DEPs unique the ADE4/CTRLE4 group.

Accession	Description	Gene Symbol	PD2.4_ADE4/CTRLE4	Scaffold_ADE4/CTRLE4	PD2.4_Abundance Ratio Adj. P-Value: ADE4/CTRLE4	Scaffold_T-Test (p-value): ADE4/CTRLE4
P01714	Immunoglobulin lambda variable 3-19	IGLV3-19	2.813	1.6	0.17	0.76
P02763	Alpha-1-acid glycoprotein 1	ORM1	2.232	1.9	0.91	0.46
P01591	Immunoglobulin J chain	IGJ; JCHAIN	1.912	2.1	0.1	0.37
P07738	Bisphosphoglycerate mutase	BPGM	1.692	1.2	0.49	0.86
P00915	Carbonic anhydrase 1	CA1	1.616	1.2	0.98	0.34
P68871	Hemoglobin subunit beta	HBB	1.607	1.5	0.34	0.49
Q13508	Ecto-ADP-ribosyltransferase 3	ART3	1.409	1.2	0.49	0.9
Q7Z7M8	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 8	B3GNT8	1.406	4.8	0.59	0.082
P80188	Neutrophil gelatinase-associated lipocalin	LCN2	1.386	1.2	1	0.75
P13716	Delta-aminolevulinic acid dehydratase	ALAD	1.317	1.8	0.15	0.61
P00751	Complement factor B	CFB	1.297	1.2	0.17	0.0098
Q92823	Neuronal cell adhesion molecule	NRCAM	1.222	1.3	0.9	0.64
P15085	Carboxypeptidase A1	CPA1	1.204	1.2	0.92	0.89
P35443	Thrombospondin-4	THBS4	0.795	0.8	0.87	0.44
Q9Y6Z7	Collectin-10	COLEC10	0.787	0.6	0.12	0.57
P15144	Aminopeptidase N	ANPEP	0.763	0.8	0.22	0.37
Q96RD9	Fc receptor-like protein 5	FCRL5	0.753	0.5	0.94	0.69

P09960	Leukotriene A-4 hydrolase	LTA4H	0.747	0.5	0.42	0.37
P52566	Rho GDP-dissociation inhibitor 2	ARHGDIB	0.718	0.1	0.037	0.13
P55072	Transitional endoplasmic reticulum ATPase	VCP	0.695	0.5	0.079	0.51
Q9Y5Y7	Lymphatic vessel endothelial hyaluronic acid receptor 1	LYVE1	0.644	0.8	0.68	0.49
Q86U17	Serpin A11	SERPINA11	0.614	0.7	0.64	0.21
P24298	Alanine aminotransferase 1	GPT	0.546	0.3	1	0.38
P35590	Tyrosine-protein kinase receptor Tie-1	TIE1	0.527	0.4	0.23	0.12
Q6UX71	Plexin domain-containing protein 2	PLXDC2	0.514	0.7	0.82	0.42
Q15555	Microtubule-associated protein RP/EB family member 2	MAPRE2	0.502	0.8	0.33	0.87
P08581	Hepatocyte growth factor receptor	MET	0.474	0.3	0.37	0.21
Q9BYE9	Cadherin-related family member 2	CDHR2	0.256	0.1	0.068	0.27
P07900	Heat shock protein HSP 90-alpha	HSP90AA1	0.591	0.4	0.22	0.61

Table 5.5. This table contains the list of 7 unique DEPs were identified in ADE4 relative to control E3. These DEPs did not clear the above criteria in ADE3/CTRLE3 or ADE4/CTRLE4 groups

Accession	Description	Gene Symbol	Abundance Ratio: (ADE4)/(CTRL E3)	Scaffold fold change_(ADE4)/(CTRLE3)	PD2.4_Abundance Ratio Adj. P-Value:ADE4/CTRLE3	Scaffold_T-Test (p-value): ADE4/CTRLE3
Q9BTY2	Plasma alpha-L-fucosidase	FUCA2	1.425	1.7	0.17	0.0098
P09172	Dopamine beta-hydroxylase	DBH	1.298	1.7	0.9	0.64
Q8I WV2	Contactin-4	CNTN4	0.790	0.8	0.92	0.89

Q12860	Contactin-1	CNTN1	0.787	0.7	0.87	0.44
Q15063	Periostin	POSTN	0.631	0.7	0.12	0.57
P22897	Macrophage mannose receptor 1	MRC1	0.659	0.7	0.22	0.37
Q16706	Alpha-mannosidase 2	MAN2A1	0.758	0.5	0.94	0.69

Table 5.6. This table contains the list of 28 DEPs were identified commonly in ADE3/CTRLE3 and ADE4/CTRLE3.

Accession	Description	Gene Symbol	Abundance Ratio: (ADE3)/(CTRL E3)	Scaffold fold change_(ADE3)/(CTRL E3)	Abundance Ratio: (ADE4)/(CTRLE 4)	Scaffold fold change_(ADE4)/(CTRL E4)	Abundance Ratio: (ADE4)/(CTRL E3)	Scaffold fold change_(ADE4)/(CTRLE3)
P35579	Myosin-9	MYH9	1.202	100	0.808	0.4	1.214	20.0
P40197	Platelet glycoprotein V	GP5	1.456	4.6	0.962	1.7	1.243	7.0
O75083	WD repeat-containing protein 1	WDR1	2.194	16	0.787	1.2	1.491	6.0
P08567	Pleckstrin	PLEK	2.438	6.8	1.456	1	2.015	4.7
P52209	6-phosphogluconate dehydrogenase, decarboxylating	PGD	1.529	5.9	1.863	0.8	1.263	4.0
P25774	Cathepsin S	CTSS	1.268	2.2	1.419	0.4	1.922	3.3
P11597	Cholesteryl ester transfer protein	CETP	1.784	3.2	0.8	1	1.607	3.2
P31146	Coronin-1A	CORO1A	2.432	11	1.352	0.3	1.776	2.0
P11279	Lysosome-associated membrane glycoprotein 1	LAMP1	1.379	2.5	1.041	1	1.349	1.8
P02788	Lactotransferrin	LTF	1.395	7.4	1.005	1.1	1.227	1.7
P04040	Catalase	CAT	1.215	1.5	0.997	1.2	1.214	1.4
P30043	Flavin reductase (NADPH)	BLVRB	1.882	1.3	1.679	0.8	1.211	1.3
P27797	Calreticulin	CALR	1.31	1.4	0.829	0.6	1.307	1.3

Q92859	Neogenin	NEO1	1.431	1.2	0.887	0.8	0.781	0.8
Q6P179	Endoplasmic reticulum aminopeptidase 2	ERAP2	0.679	0.7	0.87	0.9	0.773	0.7
P10586	Receptor-type tyrosine-protein phosphatase F	PTPRF	0.732	0.8	1.08	0.7	0.784	0.7
P07942	Laminin subunit beta-1	LAMB1	0.615	0.7	0.88	0.5	0.589	0.6
P01859	Immunoglobulin heavy constant gamma 2	IGHG2	0.667	0.3	1.245	1.1	0.733	0.5
P01861	Immunoglobulin heavy constant gamma 4	IGHG4	0.586	0.3	1.151	1.1	0.570	0.5
P08779	Keratin, type I cytoskeletal 16	KRT16	0.592	0.3	0.941	1.2	0.581	0.5
P35527	Keratin, type I cytoskeletal 9	KRT9	0.733	0.6	1.114	0.8	0.766	0.5
Q02413	Desmoglein-1	DSG1	0.471	0.5	0.69	1.1	0.493	0.4
P20930	Filaggrin	FLG	0.556	0.5	0.937	0.7	0.783	0.3
P14923	Junction plakoglobin	JUP	0.709	0.3	1.319	0.2	0.754	0.1
P07996	Thrombospondin-1	THBS1	2.142	5	0.715	0.5	1.495	2.3
O00299	Chloride intracellular channel protein 1	CLIC1	2.835	6.2	0.752	0.2	1.638	1.3
P04196	Histidine-rich glycoprotein	HRG	1.203	1.3	1.489	1.3	1.608	1.2
Q9Y490	Talin-1	TLN1	1.857	3.1	0.795	0.5	0.500	0.4

Table 5.7. This table contains the list of 9 DEPs which were identified commonly in ADE4/CTRLE4 and ADE4/CTRLE3.

Accession	Description	Gene Symbol	Abundance Ratio: (ADE3)/(CTRLE3)	Scaffold fold change_(ADE3)/(CTRLE3)	Abundance Ratio: (ADE4)/(CTRLE4)	Scaffold fold change_(ADE4)/(CTRLE4)	Abundance Ratio: (ADE4)/(CTRLE3)	Scaffold fold change_(ADE4)/(CTRLE3)
P07996	Thrombospondin-1	THBS1	2.142	5	0.715	0.5	1.495	2.3
Q13093	Platelet-activating factor acetylhydrolase	PLA2G7	1.277	0.9	1.237	1.2	1.493	1.4
Q9UNN8	Endothelial protein C receptor	PROCR	1.163	1.7	1.834	1.6	1.756	1.3
O00299	Chloride intracellular channel protein 1	CLIC1	2.835	6.2	0.752	0.2	1.638	1.3
P55103	Inhibin beta C chain	INHBC	1.231	0	1.311	1.7	1.579	1.3
P04196	Histidine-rich glycoprotein	HRG	1.203	1.3	1.489	1.3	1.608	1.2
O95479	GDH/6PGL endoplasmic bifunctional protein	H6PD	1.066	1.1	0.698	0.8	0.723	0.6
P12109	Collagen alpha-1(VI) chain	COL6A1	0.996	0.9	0.706	0.8	0.764	0.6
Q9Y490	Talin-1	TLN1	1.857	3.1	0.795	0.5	0.500	0.4

5.3.4 Changes common to both *APOE* genotypes in AD (AD risk factors independent of *APOE* allele):

DEPs common to both ADE4/CTRLE4 and ADE3/CTRLE3 groups might indicate AD pathology independent of *APOE* genotypes. The majority of DEPs were specific to each group (43 in ADE4/CTRLE4 and 87 in ADE3/CTRLE3; Figure 5.5A), while 23 DEPs were common to the ADE4 and ADE3 groups relative to their controls, as indicated in the Venn diagram (Figure 5.5A), scatter plot (Figure 5.5B) and heatmap (Figure 5.5C). Interestingly, only 9/23 DEPs were dysregulated in the same direction in both AD groups (7 elevated in the top right quadrant and 2 downregulated in the bottom left, Figure 5.5B), whereas 14/23 DEPs were dysregulated in opposite directions. The top left quadrant of Figure 5.5B shows 4/23 proteins that were elevated in ADE4/CTRLE4 but downregulated in ADE3/CTRLE3. Additionally, 10/23 proteins are shown in the bottom right, indicating downregulated proteins in ADE4/CTRLE4 but were increased in ADE3/CTRLE3. These 23 DEPs were classified into six GO-based functional groups, including metabolism and protein binding, each containing 5 DEPs (Figure 5.5E). Lipoprotein binding and cell adhesion contained 4 DEPs each. Metal ion binding and inflammation contained 3 and 4 DEPs, respectively. Some GO terms were upregulated in both AD groups, such as innate immune system, glycolysis/gluconeogenesis, myelin sheath, and the complement cascade. On the other hand, some GO terms such as lipid transport, lipid metabolism, lipoprotein metabolic process, cholesterol efflux, and focal adhesion were downregulated in ADE3/CTRLE3 and upregulated in ADE4/CTRLE4. The GO terms contain DEPs of a similar direction of fold change in both groups, suggesting disrupted pathways in AD, irrespective of *APOE* genotype.

Further, I identified 28 DEPs common in ADE4/CTRLE3 and ADE3/CTRLE3 (Figure 5.5A and Table 5.6) were plotted using a scatter plot to show the direction of fold change in common DEPs (Figure 5.5Bii). 16/28 DEPs were upregulated in both ADE4/CTRLE3 and ADE3/CTRLE3, whereas 10/28 downregulated in both the AD groups. Only 2 DEPs were dysregulated in the opposite direction, i.e., NEO1 and TLN1 were upregulated in ADE3/CTRLE3 whereas downregulated in ADE4/CTRLE3 (Figure 5.5Bii and 5.5D, Table 5.6). Next, I performed a heatmap using these 28 DEPs showing the PD2.4 abundance ratio and scaffold fold change of each protein in both AD groups (Figure 5.5D). These common DEPs were further summarized into their functional categories (Figure 5.5F and Table 5.6).

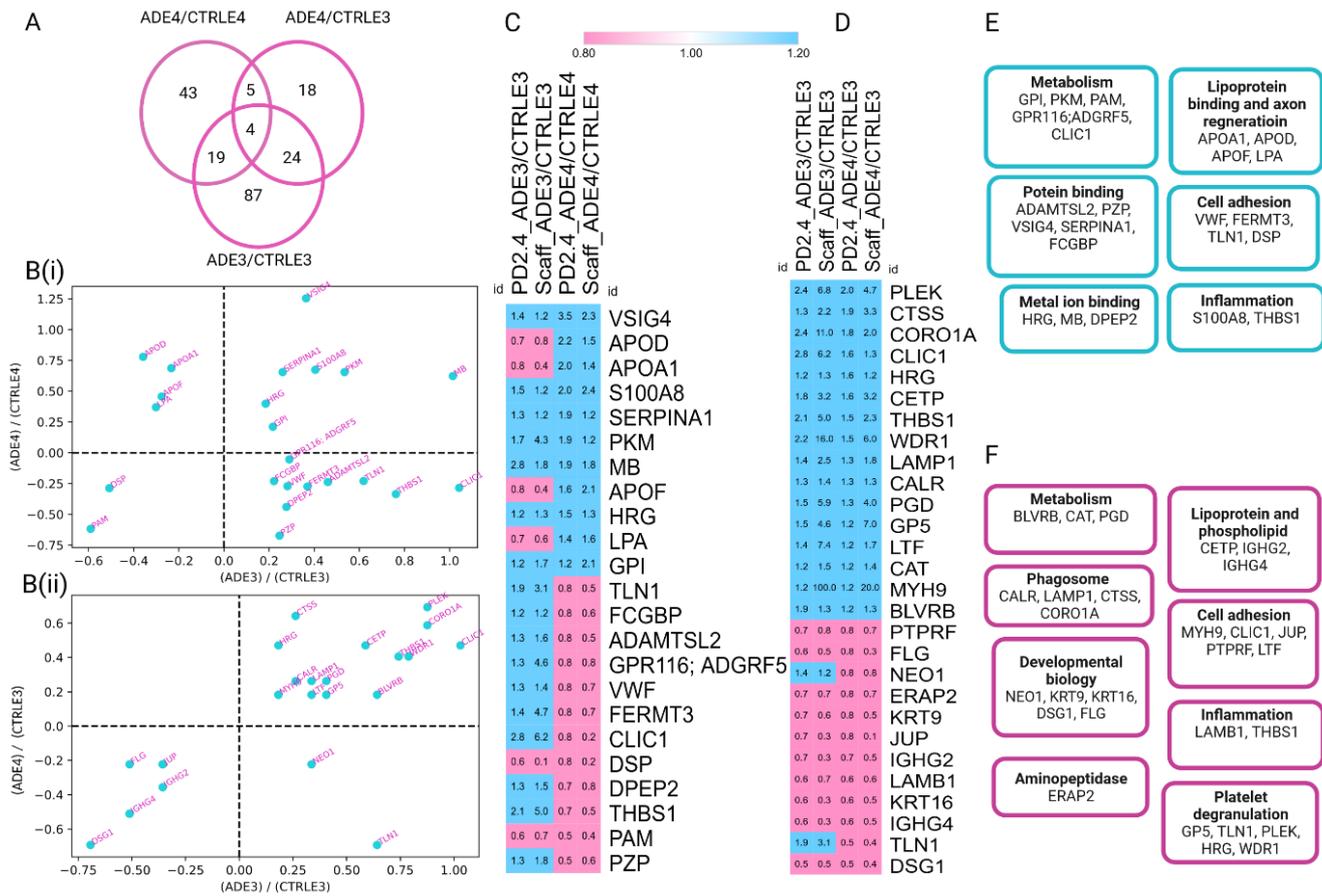


Figure 5.5: This contains common changes in ADE4 and ADE3 that might indicate the AD pathology irrespective of the presence of specific APOE genotypes. **A.** Venn diagram showing 23 DEPs common between the ADE4/CTRLE4 and ADE3/CTRLE3 groups and 28 DEPs common in ADE4/CTRLE4 and ADE3/CTRLE3 groups. **B(i).** The scatter plot shows the direction of common DEPs in ADE4/CTRLE4 and ADE3/CTRLE3 groups. These are the DEPs that vary in AD, independently of APOE allele status **C.** Heatmap using these 23 DEPs showing the PD2.4 abundance ratio and scaffold fold change of each DEPs in ADE4/CTRLE4 and ADE3/CTRLE3 groups. **D.** Heatmap using these 28 DEPs showing the PD2.4 abundance ratio and scaffold fold change of each DEPs in ADE4/CTRLE3 and ADE3/CTRLE3 groups. **E.** 23 common DEPs were broadly categorised into their biological process based on STRING software. **F.** 28 common DEPs were broadly categorised into their biological process based on STRING software.

5.3.5 Proteins linked to *APOEε4* genotype in controls:

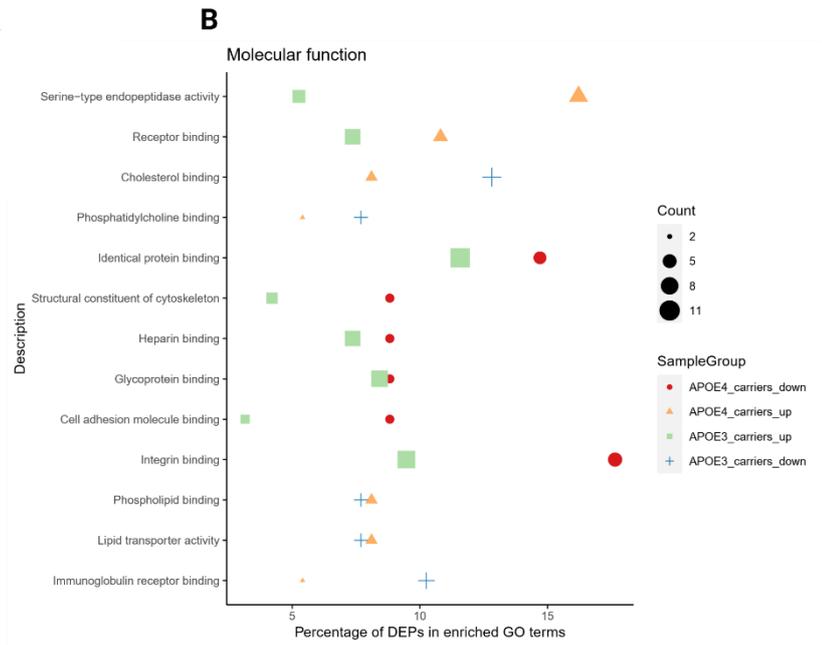
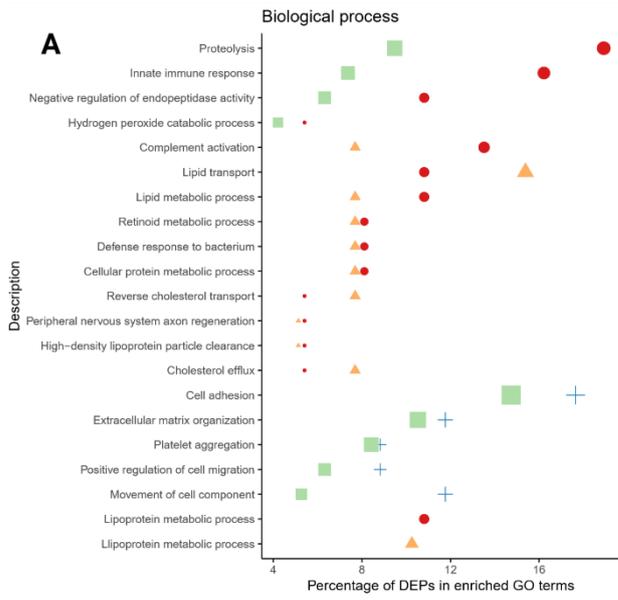
A total of 105 proteins were differentially expressed, including 48 upregulated and 56 downregulated proteins (Figure 5.1E) in control E4 relative to control E3 (CTRLE4/CTRLE3) are shown in the heatmap Figure 5.3Gi and 5.3Gii and Table 2.2. In addition, the DEPs in CTRLE4/CTRLE3 are shown in volcano plot format (Figure 5.3H).

After removing the DEPs common to both AD groups, 34 DEPs unique to CTRL E4/CTRL E3 remained, comprising 12 upregulated and 22 downregulated DEPs (Figure 5.4A and Table 5.8) (Venn diagram). This list of 34 DEPs may provide insight into potential protective mechanisms that prevented these age-matched controls from progressing to AD. These DEPs, unique to control E4 carriers, are involved in VLDL clearance (Apolipoprotein C-I, APOC1), lipid synthesis (CD5 antigen-like, CD5L), homeostasis (Complement C1q tumour necrosis factor-related protein 3, C1QTNF3 and SH3 domain-binding glutamic acid-rich-like protein 3 SH3BGRL3), which are essential in maintaining cell integrity. Additional DEPs, unique to control E4 carriers, include Fructose-bisphosphate aldolase B (ALDOB), which is associated with gluconeogenesis; signalling pathways neuropilin-2 (NRP2) and insulin-like growth factor-binding protein 7 (IGFBP7) were downregulated in the VEGFA-VEGFA2 signalling pathway while Annexin A2 (ANXA2) and ciliary neurotrophic factor receptor subunit alpha (CNTFR) are involved in Jak-STAT signalling (Figure 5.4A). As was the case with ADE4/CTRL E4 and ADE3/CTRL E3, metabolism and protein binding represent biological processes with the most significant proportion of DEPs accounting (Figure 5.4E).

Table 5.8. This table contains the list of 34 unique DEPs were identified in control E4 relative to control E3. These DEPs did not meet the above criteria in the ADE3/CTRLE3, ADE4/CTRLE3 and ADE4/CTRLE4 groups, so they are considered the DEPs unique the CTRLE4/CTRLE3 group.

Accession	Description	Gene Symbol	PD2.4_CTRLE4 /CTRLE3	Scaffold_CTRLE4 /CTRLE3)	PD2.4_Abundance Ratio Adj. P-Value:CTRLE4/CTRLE3	Scaffold_T-Test (p-value)_CTRLE4 /CTRLE3
A0A0A0MRZ8	Immunoglobulin kappa variable 3D-11	IGKV3D-11	0.48	0.09	0.01	0.10
A0A0B4J1X5	Immunoglobulin heavy variable 2-26	IGHV2-26	0.643	0.18	0.39	0.16
O14791	Apolipoprotein L1	APOL1	0.766	0.82	0.53	0.69
O43157	Plexin-B1	PLXNB1	0.802	0.78	0.62	0.58
O43866	CD5 antigen-like	CD5L	0.807	0.68	0.63	0.20
O60462	Neuropilin-2	NRP2	0.808	0.18	0.64	0.11
O94985	Calsyntenin-1	CLSTN1	0.763	0.33	0.53	0.23
P00738	Haptoglobin	HP	0.743	0.58	0.47	0.44
P01034	Cystatin-C	CST3	0.708	0.71	0.37	0.45
P01834	Immunoglobulin kappa constant	IGKC	0.45	0.34	0.00	0.03
P01876	Immunoglobulin heavy constant alpha 1	IGHA1	0.769	0.36	0.54	0.01
P02533	Keratin, type I cytoskeletal 14	KRT14	0.633	0.66	0.17	0.23
P02654	Apolipoprotein C-I	APOC1	0.678	0.15	0.39	0.16
P02753	Retinol-binding protein 4	RBP4	0.801	0.75	0.62	0.03
P05062	Fructose-bisphosphate aldolase B	ALDOB	1.729	2.08	0.15	0.17
P05362	Intercellular adhesion molecule 1	ICAM1	1.234	1.27	0.79	0.61
P06310	Immunoglobulin kappa variable 2-30	IGKV2-30	0.764	0.10	0.70	0.13

P07355	Annexin A2	ANXA2	0.732	0.50	0.63	0.59
P09211	Glutathione S-transferase P	GSTP1	1.64	3.80	0.21	0.09
P0DOX3	Immunoglobulin delta heavy chain		1.307	1.89	0.75	0.40
P0DOX5	Immunoglobulin gamma-1 heavy chain		0.64	0.45	0.19	0.06
P23141	Liver carboxylesterase 1	CES1	1.576	2.00	0.51	0.56
P24592	Insulin-like growth factor-binding protein 6	IGFBP6	0.758	0.81	0.51	0.72
P25311	Zinc-alpha-2-glycoprotein	AZGP1	0.772	0.84	0.55	0.41
P26992	Ciliary neurotrophic factor receptor subunit alpha	CNTFR	2.364	2.00	0.01	0.56
P40925	Malate dehydrogenase, cytoplasmic	MDH1	1.314	1.33	0.65	0.72
P48637	Glutathione synthetase	GSS	1.486	1.35	0.40	0.52
Q16270	Insulin-like growth factor-binding protein 7	IGFBP7	0.658	0.81	0.23	0.73
Q93063	Exostosin-2	EXT2	1.238	4.50	0.78	0.36
Q9BXJ4	Complement C1q tumor necrosis factor-related protein 3	C1QTNF3	1.421	1.35	0.50	0.66
Q9H299	SH3 domain-binding glutamic acid-rich-like protein 3	SH3BGR3	1.237	2.00	0.80	0.56
Q9NPY3	Complement component C1q receptor	CD93	1.279	1.21	0.71	0.68
Q9NSB4	Keratin, type II cuticular Hb2	KRT82	0.7	2.33	0.35	0.61
Q9UBR2	Cathepsin Z	CTSZ	0.547	0.57	0.05	0.12



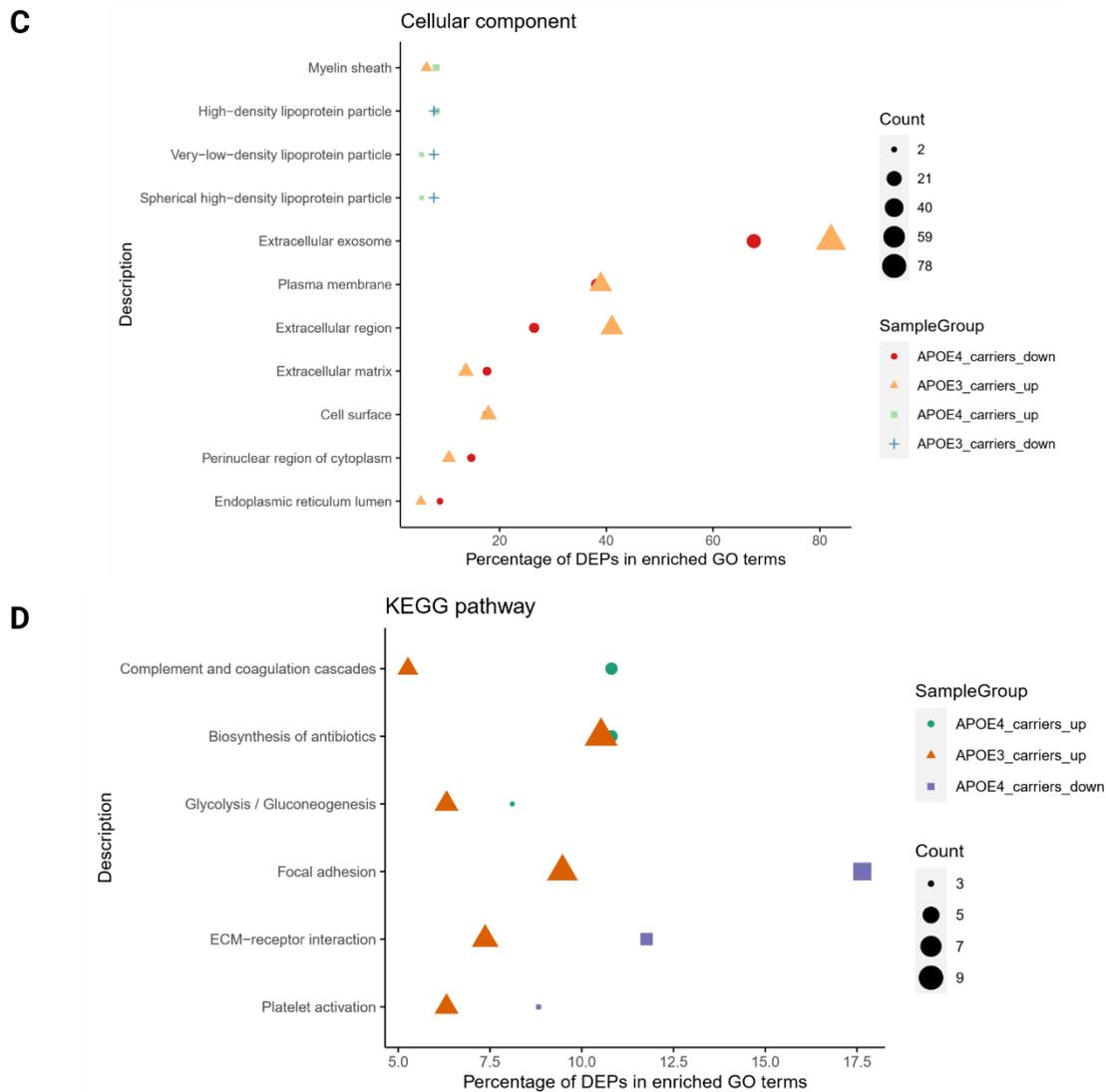


Figure 5.6: Gene ontology enrichments common to the ADE3/CTRLE3 and ADE4/CTRLE4 groups to understand the significantly dysregulated pathways. STRING software was used for enrichment analysis, and the data used in the graphical displays were prepared manually sorting from PD2.4 analyses. A. biological processes, B. molecular function, C. cellular components, and D. KEGG pathways. This figure presents a complete list of GO term enrichments, using DEPs commonly identified in both AD groups. These pathways were identified using the list of DEPs common to both ADE4 and ADE3 groups, suggesting that these pathways are disrupted in AD irrespective of the genotype. SampleGroup: Red circle- GO upregulated in ADE4; Blue plus- GO downregulated in ADE4; Green square- GO upregulated in ADE3; Yellow triangle- GO downregulated in ADE3. Count: The size of the symbols represents the number of DEPs involved in each GO enrichment term.

5.4 Discussion:

In this study, both *APOE* ϵ 3 and *APOE* ϵ 4 carriers with PiB PET imaging confirmed AD was shown to have a large number of protein expression changes in plasma, with functions including complement cascade, glycolysis, metabolism, plasma lipoprotein assembly, remodelling, and clearance. In addition, several proteins were dysregulated in the presence of the *APOE* ϵ 4 genotype relative to *APOE* ϵ 3 in both AD and control groups. This suggests that while some pathways are dysregulated by *APOE* ϵ 4, there are shared mechanisms toward developing AD independent of the *APOE* genotype. Furthermore, DEPs unique to ϵ 4 carriers in the control group suggests potential mechanisms to protect from progression to AD.

5.4.1 Plasma level of apolipoproteins and *APOE* genotype:

Apolipoproteins are among the most abundant proteins in the brain, with functions relating to cholesterol and lipid transport, and are critical for distributing and recycling lipids in the brain³³⁹. Differentially expressed apolipoproteins identified in the current study include APOA1, APOA2, APOC1, APOC3, APOD, and APOF, which were downregulated in ADE3/CTRLE3 while APOM, APOA1, APOD, and APOF were upregulated in ADE4/CTRLE4, and APOB, APOE, APOM, APOD, APOC were upregulated in ADE4/CTRLE3. Conversely, APOD, APOM, and APOA1 were downregulated in the CTRLE4/CTRLE3 group, suggesting a potentially protective effect since these same proteins are upregulated in ADE4. Notably, the level of APOF was unaffected in CTRLE4/CTRLE3, and several other apolipoproteins were downregulated in the CTRLE4/CTRLE3 group, including apolipoproteins LPA, APOL1, APOA2, APOC1, and APOC3. These opposing directions of apolipoproteins expression change in ADE4 and CTRLE4 groups suggest that rather than *APOE* ϵ 4 producing a similar "toxic effect" in both controls and AD. The response of control vs. AD subjects to the presence of the *APOE* ϵ 4 allele is qualitatively different and may be the basis of protection from disease progression in CTRLE4, while the ADE4 group succumbs to pathology. A variety of other proteins also have opposite fold change directions in the CTRLE4/CTRLE3 group compared with the ADE4/ADE3 and/or the ADE4/CTRLE4 groups. This divergent response of controls and AD subjects to the presence of the *APOE* ϵ 4 allele explains the paradoxically higher number of DEPs in the E4 "corrected" ADE4/CTRLE4 group (71 DEPs) than in the ADE4/CTRLE3 group (51 DEPs), Figures 5.1 and 5.5.

Surprisingly, by far the most extended list of DEPs was in the ADE3/CTRLE3 group (134 DEPs).

That APOC3 was downregulated in ADE3/CTRLE3 and upregulated in ADE4/CTRLE3 suggests that APOC3 may be influenced by the presence of the *APOE* ϵ 4 allele. Previous work has linked higher APOC3 levels in HDL to an increased risk of coronary heart disease and diabetes, both of which are known risk factors for dementia^{340,341}. Higher APOE levels in HDL lacking APOC3 in an elderly population were related to better cognitive function and a lower risk of AD dementia³⁴². In this context, it is of note that the lipid-binding affinity of *APOE* ϵ 4 is higher than those of *APOE* ϵ 2 and *APOE* ϵ 3^{22,23}, a property that likely accounts for the tendency of *APOE* ϵ 4 to associate with VLDL, while the E2 and E3 alleles related to HDL. Such a redistribution of lipoprotein particle composition may also affect expression, half-life, or distribution of other apolipoproteins in *APOE* ϵ 4 carriers. APOA1 is the principal structural apolipoprotein found in all HDL detectable in the blood. According to Koch et al. 2020, the presence of APOA1 in HDL does not affect the cognitive function or dementia risk, regardless of the presence of APOC3 or APOC3 in HDL³⁴³. In the current work, APOA1 was lower in ADE3/CTRLE3 but higher in ADE4/CTRLE4, suggestive of an AD-related association. Upregulation of these apolipoproteins in ADE4 may represent a homeostatic response to compensate for the deleterious *APOE* ϵ 4 allele. The functional groups involved in lipid transport, lipid metabolism, and cholesterol efflux were upregulated in ADE4, whereas all were downregulated in ADE3. Studies have suggested that cholesterol levels in the brain correlate positively with the severity of AD³⁴⁴. Elevated lipid metabolism and cholesterol efflux may be a homeostatic response facilitating cholesterol clearance in the ADE4 group³⁴⁴.

A gene ontology category enriched in all AD vs. control comparisons was metabolic changes (Figure 5.4E-H). Both ADE3 and ADE4 showed upregulation of glycolysis/gluconeogenesis-associated proteins such as glycoposphatidylinositol (GPI) and pyruvate kinase muscle (PKM). PKM catalyzes the transfer of phosphoryl groups from phosphoenolpyruvate to ADP generating ATP and pyruvate³⁴⁵. Various studies have reported that increased levels of PKM in AD CSF may indicate compensation for mitochondrial dysfunction^{346,347}. In this study, GPI and PKM were differentially expressed in both ADE3/CTRLE3 and ADE4/CTRLE4. This suggests an *APOE* allele independent effect, especially as differential expression of these two proteins was not identified in the ADE4/CTRLE3 and CTRLE4/CTRLE3 groups. Several

metabolism-related DEPs were unique to ADE3, including upregulation of 6-phosphogluconate dehydrogenase (PGD), peroxiredoxin-6 (PRDX6), isocitrate dehydrogenase, NADP (IDH1) were involved in glutathione metabolism (GSH). Not only is GSH crucial for antioxidant defence in the central nervous system, but it also plays a critical function in preserving the integrity of the blood-brain barrier³⁴⁸. As a result, alterations in GSH metabolism may have a greater impact on neurons than on other cell types³⁴⁹. However, clinical research examining the usefulness of boosting antioxidant activity in protecting or restoring cognitive functions in humans, both healthy individuals and clinical AD patients, has generally reported modest efficacy³⁵⁰. Even overexpression of the proteins involved in GSH metabolism may be insufficient to prevent/stop the damage caused by AD pathogenesis.

In the ADE4/CTRLE4 group, DEPs were identified, which were previously reported to be differentially expressed in the CSF of AD patients, including bisphosphoglycerate mutase (BPGM) carbonic anhydrase 1 (CA1) activity increased. In contrast, GSH metabolic protein, i.e., aminopeptidase N (ANPEP) activity, decreased in ADE4/CTRLE4³⁵¹. BPGM regulates the 2,3-BPG content in erythrocytes and is a critical regulator of RBC oxygen supply. Increased expression of BPGM in ADE4 implies that RBC energy enzymes are adapted to AD-related changes. Activation of the 2,3-DPG cycle results in an increase in Hb affinity for oxygen, favouring tissue hypoxia³⁵¹.

A total of 9 DEPs were identified in ADE4 compared to both control E3 and E4 suggestive of AD-related change in E4 carriers, maintained even after partial correction using E4 controls. This list includes upregulation of glucose-6-phosphate dehydrogenase (G6PD/H6PD) and platelet-activating factor acetyl-hydrolase (PLA2G7). G6PD and complementing antioxidant systems play critical roles in detoxifying reactive oxygen species (ROS). Therefore the concentration of G6PD is crucial in the antioxidant defence mechanism³⁵². A recent study by Evlice et al. 2017 reported upregulation of serum G6PD in AD *APOE* ϵ 3 carriers compared to healthy controls that might protect oxidative stress³⁵³. The downregulation of G6PD in ADE4 as compared to both control E3 and E4 carriers in the current data suggests *APOE* ϵ 4 allele-related compromise of metabolisms/antioxidant defence in AD.

Several markers related to inflammation were identified in both AD groups, including increased S100A8 expression, with the fold change being twice as large in ADE4/CTRLE4 as

in ADE3/CTRLE3. Chloride intracellular channel 1 (CLIC1) is another marker of inflammation that was found to be upregulated in ADE4/CTRLE3 and ADE3/CTRLE3 but downregulated in ADE4/CTRLE4, suggestive of an *APOE* ϵ 4 allele related to change. CLIC1 protein accumulates in peripheral blood mononuclear cells (PBMCs) and is significantly increased in the chronic inflammatory state of the CNS in neurodegenerative disease. Confocal microscopy examination and electrophysiological studies demonstrate the presence of transmembrane CLIC1 in PBMCs from Alzheimer's disease (AD) patients³⁵⁴. This enables the use of blood tests and other conventional technologies to distinguish between healthy persons and those who are undergoing neurodegenerative processes.

I found upregulation of NEO1 and NCAM1 in ADE3 carriers but no change in ADE4 compared to their respective controls. Neuronal damage markers such as Hepatocyte growth factor receptor (MET) decreased in ADE4/CTRLE4 but was not differentially expressed in ADE3/CTRLE3. The protein CHI3L1 (also called YKL40) is a well-studied CSF protein associated with reactive astrocytes, and in the current work was higher in ADE3/CTRLE3 but unchanged in ADE4/CTRLE3 and ADE4/CTRLE4³⁵⁵.

5.4.2 AD plasma proteomics in *APOE* ϵ 3 and *APOE* ϵ 4 carriers:

The *APOE* ϵ 4 allele is the most explored and familiar genetic risk factor for late-onset AD¹³, increasing the risk of AD, as well as the severity and heterogeneity of the pathology³⁵⁶⁻³⁵⁸. However, it is neither an essential nor a sufficient factor for progression to AD since non-carriers of the E4 allele also succumb to AD, while many E4 carriers do not progress to AD. Therefore, comparing AD E3 and E4 carriers with their respective E3 and E4 controls may provide insight into *APOE* allele independent proteomic associations with AD, while the same comparison using E3 controls only may provide insight into the specific contribution of the *APOE* ϵ 4 allele to the AD plasma proteome. Though it should be noted that experimental correction with normal controls who are carriers of the E3 and E4 alleles may not be perfect, since (1) the effects of *APOE* alleles may play out differently in AD vs normal controls, and (2) the E4 controls, in this case, were all heterozygous, while the AD E4 carriers were all homozygous.

The proteins PRDX2 and SOD3 are antioxidant proteins directly linked to Alzheimer's disease pathway^{335,336} and were uniquely upregulated in ADE3 compared to control E3. PRDX2, prevalent in erythrocytes, has been demonstrated to play a critical function in protecting erythrocytes from oxidative stress by scavenging ROS and contributing to cell signalling³⁵⁹. Studies have suggested that PRDX2 exists in a more oxidised state in the AD brain than controls³⁶⁰. Prx expression is increased, and the ability to retain Prxs at a decreased level is part of a unique neuroprotective process that occurs in response to A β build-up³⁶⁰. Favirn et al., 2013, investigated some consistently overexpressed genes in A β *Drosophila* (fruitflies) AD models and identified SOD3 as an A β toxicity modifier. They suggested that imbalance of this enzyme may result in an elevated level of the strong oxidant H₂O₂ in A β flies, hence contributing to AD pathology³⁶¹. The PI3K-Akt signalling pathway component collagen alpha-1(VI) (COL6A1) was decreased in ADE4 when compared to E3 and E4 controls, whereas THBS1 was decreased in ADE4/CTRLE3 and increased in ADE4 when compared E4 controls. Reducing collagen VI increased A β neurotoxicity but treating neurones with soluble collagen VI inhibited the attachment of A β oligomers with neurones, increased A β aggregation, and avoided neurotoxicity³⁶². Collagen VI is identified as a critical component of the neural damage response, and its neuroprotective potential has been demonstrated³⁶². The downregulation of these proteins uniquely in ADE4 individuals might explain the severity of the disease in *APOE* ϵ 4 carriers.

The complement system is a major part of the innate immune system, and its classical activation pathway can be directly triggered by amyloid aggregates^{363,364}. The involvement of different complement proteins in different cognitive stages suggests that triggers of the complement system may exist that are dependent on the degree of neuronal injury and/or amyloid fibril production. Previous studies have demonstrated upregulation of components of the complement system in the AD brain and the influence of the complement cascade in synapse dysfunction and loss in a mouse model of tauopathy^{365,366}. Upregulation of CFB, IGLV3-19 and downregulation of COLEC10 were uniquely identified in ADE4/CTRLE4. Comparing ADE4 with control E3 and E4, endothelial protein C receptor (PROCR) was found to be upregulated in both comparisons. Previous studies investigating complement-related protein concentrations in CSF reported divergent results with higher concentrations in AD-type dementia patients³⁶⁷⁻³⁶⁹. Notably, neuroinflammation is more severe in *APOE* ϵ 4 carriers and in related animal model studies³⁷⁰, including co-localization of *APOE* with microglia in the brain,

implying that *APOE* plays a role in the innate immune response in AD brain¹⁴⁹. Future research should focus on longitudinal changes in complement levels that occur during the development of AD and the effect of the *APOE* genotype on these processes.

5.4.3 Differential protein expression in normal controls carrying *APOE* ϵ 3 and ϵ 4 alleles:

While the *APOE* ϵ 4 allele is a well-known risk factor for AD, not all who carry this allele progress to AD, comparing the plasma proteomes of E4 and E3 carriers in normal controls may provide some insight into factors that provide protection from progression to AD despite the presence of the E4 allele. There were 14 DEPs in CTRL E4/CTRL E3 involved in metabolism, showing that dysregulation of metabolism may be a general mechanism of aging rather than a feature of AD. Glycolysis is required for a range of brain functions, including energy production, synaptic transmission, and redox balance. In both preclinical and clinical AD patients, decreased glycolytic flux has been demonstrated to correlate with the severity of amyloid and tau pathology³¹⁵. Upregulation of glycolysis/gluconeogenesis-related proteins, i.e., ALDOB and GOT1 in control E4 compared to control E3, might suggest the protective mechanism increasing the glycolysis metabolism. These metabolic changes may act as a risk indication rather than an independent risk factor. However, specific metabolism markers such as GPI and PKM may help distinguish AD from age-matched controls. A better knowledge of the link between AD and metabolism, as well as how this relationship is modulated by *APOE* ϵ 4, will also be necessary.

On the other hand, ALDOB and GOT1 might provide insights into age-matched controls' protective mechanisms. VLDL clearance, VEGFA signalling, and JAK-STAT pathways were all uniquely enriched in the case of CTRL E4/CTRL E3. Both NRP2 and IGFBP7 were downregulated in the CTRL E4/CTRL E3 group, and both are involved in VEGFA signalling. Despite the complexity and mixed evidence of VEGF associations with AD, there is growing evidence that VEGF may have a neuroprotective role³⁷¹. The VLDL clearance pathway involving APOC1 was differentially expressed in CTRL E4/CTRL E3. APOC1 is predominantly expressed in the liver and is activated during the differentiation of monocytes into macrophages required for HDL and VLDL metabolism. APOC1 has been implicated in various malignancies, and other research points to a link between APOC1 and human

longevity^{372,373}. Given the discrepancy of research findings, it is critical to discover the role of these pathways in human longevity and healthy aging.

4.5 Conclusion

This study performed an in-depth proteome analysis to identify plasma proteome signatures associated with *APOE* ϵ 3 and *APOE* ϵ 4. In late-onset AD, the *APOE* ϵ 4 allele is the most well-known genetic risk factor. However, non-carriers of the E4 allele also succumb to AD, but many E4 carriers do not. We identified a high number of protein expression alterations in plasma which were found uniquely in *APOE* ϵ 3 and *APOE* ϵ 4 carriers. Interestingly, several proteins were also dysregulated in the presence of both *APOE* ϵ 3 and *APOE* ϵ 4 genotypes depicting the involvement of these proteins in the pathogenesis of AD regardless of the *APOE* genotypes. Furthermore, our findings also identified some proteins previously discovered in AD CSF and brain proteomics signatures that could provide clinically meaningful information.

Chapter 6

General Discussion and Future Directions

6.1 Discussion and conclusions:

Alzheimer's disease (AD) is a multifactorial neurodegenerative disease characterised by various cellular and molecular processes dysregulation. Most studies focus on A β and tau, which are the primary components of amyloid plaques and neurofibrillary tangles, respectively. However, the search for a disease-modifying medication has yet to yield clinically meaningful results, and it is becoming increasingly apparent that dementia involves a plethora of other pathological changes across various pathways. Proteomic investigations based on mass spectrometry (MS) provide an efficient and comprehensive method for examining hundreds of proteins quantitatively simultaneously, utilising small amounts of biofluids. Previous proteomics studies have led to the discovery and identification of thousands of potential biomarkers for cancer and other diseases^{374,375}.

In chapter 2, the meta-analysis suggested a relative paucity of additional biomarkers and low numbers of reported studies in AD was notable, in particular, those using MS-based proteomics. Apart from core biomarkers (A β and tau), only a handful of additional markers of neurodegeneration and inflammation, such as CSF NfL, IgG, and IL-6, were replicated in previous studies. I showed that the range of biomarkers studied had been heavily focused on CSF, but for establishing promising blood tests, more research is needed before any of these could complement CSF or imaging tests for non-invasive testing. Human blood is an often-used clinical sample for diagnostic analysis because it is less invasive, currently in routine clinical usage, and comes into direct contact with damaged organs/tissues, frequently reflecting the downstream effects of disease processes. Therefore, I argued in this chapter that an in-depth investigation of the plasma proteome might reveal additional possible markers that might be used and tested for their utility in clinical AD diagnosis. However, plasma proteomics has historically been limited by the lack of throughput and sensitivity, owing mainly to the complexity of the plasma proteome estimated to be between 9 and 13 orders of magnitude between the lowest and highest abundance protein.

Chapter 3 began with the attempt to overcome this dynamic range limitation by comparatively evaluating several fractionation approaches that might facilitate greater plasma proteome coverage and identify tissue-specific proteins, while remaining compatible with the greater sample throughput required by clinical research. In chapter 3, I evaluated eight relatively simple prefractionation methods for plasma high-abundance protein depletion and low-

abundance protein enrichment to significantly increase the coverage of the typical plasma proteome. Using a two-step approach, (1) I compared high-abundance protein depletion (immunoaffinity columns; Hu6, Hu14) and low-abundance protein enrichment (ProteoMiner combinatorial peptide ligands), followed by (2) further fractionation of the low-abundance proteins using chromatographic and electrophoretic approaches (C18 column and one-dimensional (1D) sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS PAGE), respectively). A total of 4385 proteins were identified with high confidence using LC-MS/MS (data-dependent analysis) and database searching with MaxQuant and Trans-Proteomic Pipeline. Of the total 4385, 3064 (almost 70%) proteins were common across all methods, and high correlations were seen for all within-method technical replicates, as well as reasonable correlations for between-method technical duplicates. Along with great coverage, our plasma proteomics data revealed proteins from 51 organs/tissues systems of the human body using the human protein atlas software, demonstrating that plasma functions as a fluid that reflects the unique protein fingerprints of each organ. The work presented here illustrates that more straightforward and faster procedures can achieve comparable levels of proteome coverage. Furthermore, these approaches are accessible to the majority of basic biochemistry laboratories.

Several possible biomarkers were discovered in our data that are typically present in plasma at ng/mL levels, including superoxide dismutase (SOD2), ribonuclease 4 (RAB4), 72 kDa type IV collagenase matrix metalloproteinase 2 (MMP2), and serpin (SERPIN) proteins. These proteins have been implicated in disease, including RAB4, which protects neurones from degeneration in amyotrophic lateral sclerosis (ALS) by increasing neurofilament production and protects mouse embryonic cortical neurones from hypothermia-induced degeneration. Another protein identified in our data collection is matrix metalloproteinase 2 (MMP2). This enzyme is implicated in vascular remodelling, neural progenitor cell migration, and tumour invasion. In addition, I detected many low-abundance plasma proteins, including macrophage colony-stimulating factor 1 (200 pg/mL), P-selectin (120 pg/mL), platelet-derived growth factor receptor (3 pg/mL), and leptin (4.7 ng/mL), several of which have been previously reported.

Our results also demonstrated the reproducibility of these prefractionation techniques and confirmed that gel-based procedures can be used in place of expensive and time-consuming chromatographic column separation, considerably shortening the time required for analysis and cost-effectiveness. Consequently, I observe that all eight methods perform well in identifying

high numbers of proteins and achieving comparable protein identifications to the maximum proteome identified in human plasma. Furthermore, my findings expanded the repository of plasma proteome, which is not limited to specific diseases but may be used to investigate any pathophysiological situation in plasma. In other words, it is possible to conduct a targeted proteomics experiment and begin validating the proteins using this spectrum library. In conclusion, this chapter has demonstrated that various methodologies allow for customization of workflows based on project-specific requirements, such as whether qualitative or quantitative information is required, the size of the project in terms of sample numbers, and the availability of specific laboratory resources.

Our meta-analysis (chapter 2) demonstrated that only a few proteomics studies had been established to dissect molecular pathways of mild cognitive impairment (MCI) and dementia pathology, especially in longitudinal cohorts. Longitudinal data serve an important role in understanding the pathological changes in AD, as some studies suggest AD pathology might start 15-20 years prior to symptom onset^{180,376}. After successfully covering the in-depth plasma proteome (chapter 3), the major focus of chapter 4 was to provide insights into the differential expression (up and down-regulation) of proteins that are affected in normal ageing, MCI and AD in both longitudinal and cross-sectional analyses. In chapter 4, I selected and applied the method described in the chapter 3, using a two-step plasma fractionation approach; HU14 removal of high abundance plasma proteins, followed by parallel 1D SDS/PAGE of the low abundance protein fractions for the proteome profiling of clinical samples.

I have begun to address two major obstacles in identifying plasma protein biomarkers for the common age-related neurodegenerative diseases in this work: (1) the restricted current level of information regarding the longitudinal changes in the plasma proteome in normal vs diseased individuals, and (2) the even more limited knowledge of the preclinical AD plasma proteome. In chapter 4, plasma samples were analysed from participants with plasma samples from two time-points; baseline (Wave 1) and 6 years period of follow up (Wave 4), with 11 participants remaining normal, and the remainder progressing to mild cognitive impairment (MCI) and AD (n=11 each) from the Sydney Memory and Ageing Study (MAS) cohort. Extensive longitudinal and cross-sectional proteomics data was obtained to facilitate several hypothesis-driven studies of AD to attempt to fill knowledge gaps regarding the progression of AD. Expression changes were observed in a surprisingly large number of plasma proteins (71 DEPs) in normal ageing (6 years follow up period). The 71 DEPs identified were involved in three major functional

groups, with the highest number of age-related DEPs being in cell signalling (35%), cytoskeleton and microtubules function (17%), and metabolism (15%). Several other categories represented $\leq 8\%$ of total DEPs each. The Hippo signalling pathway was particularly enriched with ageing. This signalling pathway included DEPs of the 14-3-3 protein family (YWHAZ, YWHAH, YWHAE, YWHAB, YWHAQ, YWHAB) and actin Gamma 1 (ACTG1), which were all upregulated in the longitudinal data. The 14-3-3 protein family is highly expressed in the brain and influences many aspects of brain function through interactions with a diverse set of binding partners involved in neural signalling, neuronal development, and neuroprotection¹⁷⁰, and is a well-studied protein family in AD CSF^{171,172}. Longitudinal analysis shows that altered plasma expression of the 14-3-3 protein family is an age-related change, being observed in all three longitudinal analysis groups (cognitively normal controls, MCI and AD), so it may have functional implications for progression to MCI and/or AD since ageing is the major risk factor for these conditions²⁶⁸.

After excluding ageing-related proteins (71 DEPs) in the longitudinal analysis of MCI and AD, I identified a long list of DEPs (66 and 60 DEPs specific to MCI and AD, respectively), with similarly significant numbers in the cross-sectional analyses (89 and 70 MCI and AD specific DEPs respectively). Such a high number of DEPs raises the question of which DEPs might be suitable biomarker candidates. To choose a potential list of biomarkers, I applied two strategies as follows:

1. To select only those proteins with robust expression change between groups, I used the following inclusion criteria: proteins quantified in >6 individuals, proteins identified with a minimum of two peptides per protein, a consistent direction of protein fold change across two bioinformatics platforms with orthogonal quantification approaches (peak area ratio with PD2.4 and spectral counting with Scaffold) with a fold change of at least 20% (≤ 0.08 and ≥ 1.2) in both search engines. These are orthogonal approaches, each with specific advantages and disadvantages^{265,266}, so I reasoned that the most reliable changes should be consistent across both search engines, i.e., PD2.4 and Scaffold.

2. A list of proteins was selected based on the consistency of fold-change in longitudinal and cross-sectional analyses for AD and MCI groups.

I identified 16 and 15 common DEPs in longitudinal and cross-sectional MCI and AD, respectively. Of the 15, S100A7, PAM, ALDOB, FAM3C, VCP, and TF are some potential

markers of dementia and probable AD reported in the previous literature as they are related to AD pathology. Of these proteins, FAM3C and MAN2A2 are involved in metabolism; ALDOB and VCP in glycolysis; PAM in the post-translational modification; and S100A7 and TF in the immune system. A noteworthy feature of our study is the identification of brain-derived proteins such as MAN2A2, PAM, TF, QDPR, FAM3C, which have previously been reported to be dysregulated in AD CSF and brain³⁰⁴.

In MCI, LTF, PSME1, HSPA4, COL5A1, ASGR2, and PPP1R7 comprise a significantly shorter list, but they contain considerably more promising candidates for further validation work. MCI is frequently regarded as a risk factor for and/or prodromal stage of AD (thus, it was noteworthy that in longitudinal and cross-sectional analyses, only roughly 18% (19 proteins) and 13% (18 proteins) of DEPs were common to both AD, and MCI, respectively. In this respect, it is worth noting that the overwhelming majority of DEPs found in AD, and MCI are condition-specific rather than shared.

As stated previously, the preclinical stage of neurodegeneration is an appealing target for disease-modifying intervention in AD^{142,143}. The considerably larger pool of preclinical AD-associated DEPs may in part be evidence of pathology in progress, in addition to providing a large pool of putative early biomarkers (ADW1/CTRLW1). With the reasoning that the most robust biomarkers may continue to be observed with clinical disease onset, 15 DEPs were shared with clinical AD (ADW4/CTRLW4). Apart from the well-known CSF AD biomarker (CHI3L1), I propose a list of novel markers, including PSMB2, PAM, ALDOB, TF, MAPRE2, and VCP, which may be possible AD preclinical biomarkers, being dysregulated in all three AD comparison groups, i.e., longitudinal (ADW4/ADW1), incipient AD (ADW4/CTRLW4) and preclinical AD (ADW1/CTRLW1). These DEPs are involved in metabolism, glycolysis, immune system, and proteasome, dysregulated in preclinical AD stage.

An interesting finding is that most DEPs are upregulated (~66 % each) in the incipient MCI and AD groups. In contrast, DEPs with consistent fold changes are predominantly downregulated in the preclinical and incipient AD groups, with only two DEPs (< 20%) being upregulated. Interestingly, glycolysis and gluconeogenesis presented as top GO terms with significant enrichment in preclinical and clinical AD plasma in this unbiased analysis. This concurs with three glycolytic proteins from our 8-protein signature being PGK1, VCP and ALDOB. Various studies have demonstrated that dysregulation of glucose metabolism in the

brain is a prominent abnormality during the preclinical stage of AD^{315,316,317}. This suggests that the glycolysis-related protein dysfunction associated with AD may result in bioenergetic and biosynthetic disturbances, disrupting metabolic and synaptic homeostasis and resulting in abnormal protein deposition and progression to dementia.

In chapter 2, the meta-analysis showed that apart from environmental and lifestyle effects, genetic variation of the apolipoprotein E (*APOE*) gene plays a significant role in AD risk and progression. However, the impact of genetic variation is poorly understood from a mechanistic perspective. Recent studies have reported that approximately 65% of individuals with late-onset familial and sporadic AD bear the *APOE* ϵ 4 allele⁴. One copy of *APOE* ϵ 4 is associated with a threefold increase in the disease risk, while two copies are associated with a more than tenfold increase in risk¹⁵. Previous studies indicate a dose-dependent effect of *APOE* ϵ 4 on CSF A β ₄₂ level in LOAD. However, insufficient studies were available to explore the difference between *APOE* ϵ 3 and *APOE* ϵ 4 dose effects on AD plasma proteome.

In chapter 5, the relationships between the plasma proteome, *APOE* genotypes and AD pathology were studied. Therefore, I have profiled an in-depth plasma proteome to understand the impact of *APOE* ϵ 3 and ϵ 4 carriage in AD dementia (verified with PiB PET neuropathology imaging) relative to *APOE* ϵ 3 and ϵ 4 control carriers with normal cognition. In turn, I identified 23 DEPs, including S100A8, PAM, CLIC1, APOD and APOA1, commonly dysregulated in the presence of *APOE* ϵ 3 and *APOE* ϵ 4 relative to their respective controls. These DEPs may be involved in developing AD independent of the *APOE* genotype. The common 23 DEPs involved molecular pathways such as metabolism, inflammation and plasma lipoprotein assembly, remodelling, and clearance that are shared mechanisms toward AD development. This chapter shows 65 DEPs (48 upregulated and 17 downregulated) proteins that were dysregulated exclusively in AD *APOE* ϵ 3 relative to control *APOE* ϵ 3 (Table 5.2). Only 29 DEPs (13 were upregulated and 16 downregulated) were dysregulated exclusively in AD *APOE* ϵ 4 relative to control *APOE* ϵ 4. The DEPs were manually categorized using the PD2.4 analyses; most proteins are involved in metabolism (34%), protein binding (21%) and signalling (18%). Other enriched categories were complement cascade, PI3K-Akt pathway, post-translational modification (PTM), protein digestion and protease inhibition. The PI3K-Akt pathway, which includes heat shock protein HSP 90-alpha (HSP90AA1), collagen alpha-

1(VI) chain (COL6A1), and thrombospondin-4 (THBS4), was one of the distinct pathways dysregulated.

Seven proteins were upregulated, including FUCA2 and DBH, whereas there was downregulation of CNTN4, CNTN1, MAN2A1, MRC1 and POSTN in the AD *APOE* ϵ 4 relative to control *APOE* ϵ 3. As these proteins are uniquely expressed in *APOE* ϵ 4 +ve individuals, they most likely reflect the effects of the *APOE* ϵ 4 allele on AD.

Next, I identified 44 DEPs in both AD *APOE* ϵ 3 and AD *APOE* ϵ 4, replicated in our previous cross sectional and longitudinal AD MAS cohort from chapter 4 (Table 6.1). Plasma proteins that had changed in the same direction in all four AD comparisons included myoglobin (MB), mannose-binding protein (MBL2), prostaglandin-H2 D-isomerase (PTGDS), and phosphatidylethanolamine-binding protein 4 (PEBP4). In addition, proteins replicated in age-matched AD and control group (cross-sectional AD) and *APOE* ϵ 4 included upregulation of apolipoprotein M (APOM), apolipoprotein A-I (APOA1), and downregulation of transitional endoplasmic reticulum ATPase (VCP), microtubule-associated protein RP/EB family member 2 (MAPRE2), fermitin family homolog 3 (FERMT3) and others. Tenascin (TNC), chitinase-3-like protein 1 (CHI3L1) were upregulated, while junction plakoglobin (JUP), Immunoglobulin heavy constant mu (IGHM) were downregulated in both AD *APOE* ϵ 3 and cross-sectional AD replication cohorts.

Table 6.1. This table contains the list of 44 DEPs common between AD APOEε3 and AD APOEε4 (AIBL cohort) and cross sectional and longitudinal AD (MAS cohort).

Accession	Description	Gene Symbol	PD2.4_Abundance Ratio: (AD, E3) / (CTRL, E3)	Scaffold_Fold Change by Category_AD E3CTRL E3	PD2.4_Abu ndance Ratio: (AD, E4) / (CTRL, E4)	Scaffold_F old Change by Category_ ADE4CTR LE4	PD2.4_Abundance Ratio: (ADW4) / (ADW1) PD2.4	Scaffold_Fol d change (ADW4) / (ADW1)_ Scaffold	PD2.4_Abu ndance Ratio: (ADW4) / (CTRLW4)	Scaffold fold change (ADW4) / (CTRLW4)
O95445	Apolipoprotein M	APOM	0.968	0.7	2.457	2.3	1.476	1.3	1.506	1.4
P01024	Complement C3	C3	0.818	0.7	1.211	1.4	0.803	0.7	1.022	1
P01861	Immunoglobulin heavy constant gamma 4	IGHG4	0.586	0.3	1.151	1.1	0.757	0.2	0.597	3.3
P01871	Immunoglobulin heavy constant mu	IGHM	0.461	0.5	0.883	1.2	0.827	0.8	0.597	0.6
P02144	Myoglobin	MB	2.757	1.8	1.864	1.8	1.244	0.9	2.961	3.5
P02647	Apolipoprotein A-I	APOA1	0.792	0.4	1.987	1.4	1.355	1.4	1.49	1.2
P04083	Annexin A1	ANXA1	2.394	7	4.754	0.8	1.398	3.9	0.745	0.7
P04406	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	1.44	1.3	1.512	0.6	1.536	2.2	0.952	0.8
P05090	Apolipoprotein D	APOD	0.699	0.8	2.179	1.5	1.247	1.1	1.384	1.4
P05164	Myeloperoxidase	MPO	1.485	4.7	1.042	0	1.573	1.3	2.039	3.3
P06733	Alpha-enolase	ENO1	1.89	3.5	1.168	0.4	1.27	2.7	0.652	0.7
P06744	Glucose-6-phosphate isomerase	GPI	1.242	1.7	1.235	2.1	1.215	2	0.778	0.4
P07996	Thrombospondin-1	THBS1	2.142	5	0.715	0.5	1.187	2.2	0.713	0.6
P08294	Extracellular superoxide dismutase [Cu-Zn]	SOD3	1.669	1.4	1.083	0.8	1.462	1.2	1.078	1
P08519	Apolipoprotein(a)	LPA	0.74	0.6	1.447	1.6	1.15	0.8	0.54	0.5
P08567	Pleckstrin	PLEK	2.438	6.8	1.456	1	1.123	2.3	0.667	0.7

P11226	Mannose-binding protein C	MBL2	1.182	1.2	1.398	1.2	1.215	1.3	1.456	1.4
P12814	Alpha-actinin-1	ACTN1	1.726	5.2	0.949	1.1	1.207	2.1	0.643	0.6
P14618	Pyruvate kinase PKM	PKM	1.706	4.3	1.925	1.2	1.442	3.2	0.49	0.5
P14923	Junction plakoglobin	JUP	0.709	0.3	1.319	0.2	0.737	0.3	0.689	0.4
P15085	Carboxypeptidase A1	CPA1	3.117	0.9	1.204	1.2	1.125	1.7	0.812	0.9
P18206	Vinculin	VCL	1.348	1.9	0.801	0.5	1.501	1.6	0.779	0.6
P23284	Peptidyl-prolyl cis-trans isomerase B	PPIB	3.097	1.5	1.011	0.2	1.536	3.4	1.142	0.9
P24821	Tenascin	TNC	1.24	1.5	1.065	1	1.21	0.9	1.481	1.3
P25774	Cathepsin S	CTSS	1.268	2.2	1.419	0.4	0.877	0.8	1.299	1.1
P26038	Moesin	MSN	1.008	1.8	0.704	0.7	1.307	1.7	0.674	0.7
P27797	Calreticulin	CALR	1.31	1.4	0.829	0.6	1.485	2.3	0.928	0.9
P30043	Flavin reductase (NADPH)	BLVRB	1.882	1.3	1.679	0.8	0.927	1	1.917	2.1
P31146	Coronin-1A	CORO1A	2.432	11	1.352	0.3	1.298	5.6	0.664	0.6
P36222	Chitinase-3-like protein 1	CHI3L1	2.318	2.1	1.363	0.6	0.675	1.2	1.242	1.6
P41222	Prostaglandin-H2 D-isomerase	PTGDS	1.259	0.9	1.257	1.5	1.29	1.4	1.65	1.5
P52209	6-phosphogluconate dehydrogenase, decarboxylating	PGD	1.529	5.9	1.863	0.8	1.237	2.1	0.765	0.7
P55072	Transitional endoplasmic reticulum ATPase	VCP	1.213	INF	0.695	0.5	1.347	1.8	0.704	0.5
P55103	Inhibin beta C chain	INHBC	1.231	0	1.311	1.7	1.254	1.8	1.533	1.3
P61981	14-3-3 protein gamma	YWHAG	1.621	INF	0.502		1.452	2.5	0.849	0.9
P80188	Neutrophil gelatinase-associated lipocalin	LCN2	2.625	1	1.386	1.2	1.555	1.3	1.648	1.2

P81605	Dermcidin	DCD	2.173	1.8	1.321	INF	0.395	0.6	0.787	0.6
Q12907	Vesicular integral-membrane protein VIP36	LMAN2	1.447	2.8	1.292	0.2	1.212	2	0.991	1.5
Q15555	Microtubule-associated protein family RP/EB member 2	MAPRE2	1.753	INF	0.502	0.8	1.639	3.7	0.598	0.3
Q86UX7	Fermitin family homolog 3	FERMT3	1.448	4.7	0.76	0.7	1.338	3.2	0.758	0.6
Q96S96	Phosphatidylethanolamine-binding protein 4	PEBP4	1.352	1.3	1.48	1	1.37	1.2	1.061	1.2
Q9H299	SH3 domain-binding glutamic acid-rich-like protein 3	SH3BGL3	3.244	1.3	0.266	0	1.77	2.5	1.279	1.1
Q9Y490	Talin-1	TLN1	1.857	3.1	0.795	0.5	1.503	2.3	0.659	0.8

While these findings require confirmation in additional independent disease cohorts, this investigation has identified candidate proteins that can be used to identify individuals at risk of cognitive decline and eventually developing dementia. The findings shed light on a set of markers not previously associated with MCI or AD, and point the way forwards for study of the biological pathways that may be targeted for therapies to reduce the progression of cognitive decline. Interestingly, global proteomics analysis revealed molecular similarities and differences at different stages of dementia. The extensive longitudinal and cross-sectional data facilitate several AD hypothesis-driven studies, elucidating knowledge gaps between preclinical and clinical stages of the illness. Numerous possible indicators that alter in a similar direction to prior CSF and brain samples improve our mechanistic understanding of AD aetiology and encourage the development of novel strategies for diagnosing and treating this debilitating disease. Blood-based biomarkers have the potential advantage of being more accessible than practically any other body fluid, and blood-based diagnostics lend themselves to high-throughput and low-cost assessments. Plasma proteomics can now detect proteins derived from the brain and generate consistent protein signatures across many independent search engines, paving the path for future research on biomarker identification in neurodegenerative disorders. Additionally, I expect that the approach outlined here would identify additional clinically and etiologically significant plasma biomarkers.

6.2 Limitations:

The work presented in this thesis has a number of limitations: 1). The overall sample number was relatively modest, reducing the power of the analysis. 2). A relative quantification approach was used, so absolute level of expression change requires use of targeted approaches 3). Only two independent replication cohorts were used. 4). Individuals with AD were all homozygous *APOE* $\epsilon 4/4$ carriers, whereas age-matched controls were heterozygous *APOE* $\epsilon 3/4$ carriers. These limitations were discussed in greater detail in the respective chapters. Therefore, the work should be regarded as exploratory, and additional research into the relevance of these proteins is warranted in prospective studies of dementia-free individuals, preferably beginning in mid-life and followed up until the development of dementia.

6.3 Future studies

Even though it is now possible to identify over 4500 proteins in plasma using advanced LC-MS/MS and data analytics, these approaches generally rely on complex workflows that include depletion, protein fractionation, peptide fractionation, and isobaric labelling coupled to LC-MS/MS, which is time-consuming (days to weeks) and forces a trade-off between depth of protein coverage and sample throughput. These constraints impede the development of novel protein-based disease biomarkers and act

as roadblocks to the widespread adoption of proteogenomics and the annotation of genomic variations by proteins in the future.

In concept, mass spectrometry (MS)-based proteomics is ideally suited for doing effective protein biomarker discovery investigations with large enough sample cohorts. Over the past many years, the technical capability of mass spectrometry-based proteomics has increased dramatically. Currently, the MS biomarker discovery technique offers a number of important characteristics that might make it the best platform for protein biomarker discovery. These characteristics include high specificity, extensive protein coverage, and the accessibility of PTMs. The use of data-dependent acquisition (DDA) mode also results in sparse datasets with an irreproducible sampling of the peptides in the sample, making statistical analysis more complex and the power of high sample numbers reduced. Identifying proteins in large clinical cohorts as quickly and consistently as possible is a crucial objective for clinical proteomics. As proteomics pipelines develop in terms of throughput and depth, at least two strategies are being pursued: (1) employing advanced acquisition modes, such as BoxCar or scanning SWATH; or state-of-the-art LC/MS setups, such as ion mobility-enabled PASEF; and (2) improving sample preparation, either by enriching low-abundant proteins or by using more sophisticated data processing pipelines that leverage additional information across and within samples; and (3) improving sample preparation. These two tactics are often used in conjunction with one another to improve performance. Approaches that improve proteome coverage via sample preparation (Strategy 2) often result in a more complicated and less scalable procedure, despite advancements in, and especially when paired with, sample preparation automation.

As the throughput, depth, and robustness of mass spectrometry (MS)-based proteomics continue to improve, I anticipate that it will be more extensively used in routine clinical practice. For example, MS-based proteomics will be used to evaluate biomarker panels rather than single markers since it takes advantage of the fact that it is intrinsically multiplexed and allows for a more thorough characterization of clinical situations than single markers. In addition, these biomarker panels might be regularly supplemented with high-quality marker panels, such as those described here, to aid in the development of biomarker-guided choices in patients with clinical AD.

In summary, this thesis has provided a wealth of data for further study and the generation of hypotheses. I have added substantial new knowledge on proteins and pathways involved in AD pathogenesis from a broad to specific biomarkers perspective. Future studies of the reported pathways could elucidate the involvement of specific proteins in MCI and dementia due to AD. The next step is to choose a targeted biomarker from 44 common DEPs in MAS and AIBL cohorts (Table 6.1) to be confirmed as an AD biomarker using targeted proteomics on the Dominantly Inherited Alzheimer Network (DIAN) cohort. As this work is exploratory, additional research into the relevance of these proteins should be carried out in prospective studies of dementia-free persons in midlife with long-term follow-up to incident dementia.

References

1. Przedborski S, Vila M, Jackson-Lewis V. Neurodegeneration: what is it and where are we? *The Journal of clinical investigation*. 2003;111(1):3-10.
2. Höhn A, Tramutola A, Cascella R. Proteostasis Failure in Neurodegenerative Diseases: Focus on Oxidative Stress. *Oxid Med Cell Longev*. 2020;2020:5497046-5497046.
3. Dugger BN, Dickson DW. Pathology of Neurodegenerative Diseases. *Cold Spring Harbor perspectives in biology*. 2017;9(7):a028035.
4. Bekris LM, Yu C-E, Bird TD, Tsuang DW. Genetics of Alzheimer disease. *J Geriatr Psychiatry Neurol*. 2010;23(4):213-227.
5. Raz L, Knoefel J, Bhaskar K. The neuropathology and cerebrovascular mechanisms of dementia. *J Cereb Blood Flow Metab*. 2016;36(1):172-186.
6. 2020 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2020.
7. Prince MJ, Wu F, Guo Y, et al. The burden of disease in older people and implications for health policy and practice. *Lancet*. 2015;385(9967):549-562.
8. Mendez MF. Early-Onset Alzheimer Disease. *Neurol Clin*. 2017;35(2):263-281.
9. Vandembroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology*. 2007;18(6):805-835.
10. McDade E, Wang G, Gordon BA, et al. Longitudinal cognitive and biomarker changes in dominantly inherited Alzheimer disease. *Neurology*. 2018;91(14):e1295-e1306.
11. Suárez-Calvet M, Araque Caballero MÁ, Kleinberger G, et al. Early changes in CSF sTREM2 in dominantly inherited Alzheimer's disease occur after amyloid deposition and neuronal injury. *Sci Transl Med*. 2016;8(369):369ra178-369ra178.
12. Fagan AM, Xiong C, Jasielec MS, et al. Longitudinal change in CSF biomarkers in autosomal-dominant Alzheimer's disease. *Sci Transl Med*. 2014;6(226):226ra230.
13. Liu C-C, Liu C-C, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol*. 2013;9(2):106-118.
14. Farrer LA, Cupples LA, Haines JL, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *Jama*. 1997;278(16):1349-1356.
15. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*. 1993;261(5123):921.
16. Chow VW, Mattson MP, Wong PC, Gleichmann M. An overview of APP processing enzymes and products. *Neuromolecular Med*. 2010;12(1):1-12.

References

17. Dahlgren KN, Manelli AM, Stine WB, Jr., Baker LK, Krafft GA, LaDu MJ. Oligomeric and fibrillar species of amyloid-beta peptides differentially affect neuronal viability. *J Biol Chem.* 2002;277(35):32046-32053.
18. Mattson MP. Pathways towards and away from Alzheimer's disease. *Nature.* 2004;430(7000):631-639.
19. Schupf N, Tang MX, Fukuyama H, et al. Peripheral Abeta subspecies as risk biomarkers of Alzheimer's disease. *Proc Natl Acad Sci U S A.* 2008;105(37):14052-14057.
20. Atwood CS, Scarpa RC, Huang X, et al. Characterization of copper interactions with alzheimer amyloid beta peptides: identification of an attomolar-affinity copper binding site on amyloid beta1-42. *J Neurochem.* 2000;75(3):1219-1233.
21. Olsson B, Lautner R, Andreasson U, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol.* 2016;15(7):673-684.
22. Saido T, Leissring MA. Proteolytic degradation of amyloid β -protein. *Cold Spring Harb Perspect Med.* 2012;2(6):a006379.
23. Bateman RJ, Munsell LY, Morris JC, Swarm R, Yarasheski KE, Holtzman DM. Human amyloid-beta synthesis and clearance rates as measured in cerebrospinal fluid in vivo. *Nat Med.* 2006;12(7):856-861.
24. Mawuenyega KG, Sigurdson W, Ovod V, et al. Decreased clearance of CNS beta-amyloid in Alzheimer's disease. *Science.* 2010;330(6012):1774.
25. Iqbal K, Liu F, Gong CX, Grundke-Iqbal I. Tau in Alzheimer disease and related tauopathies. *Curr Alzheimer Res.* 2010;7(8):656-664.
26. Iqbal K, Alonso A, Gong C, et al. Molecular pathology of Alzheimer neurofibrillary degeneration. *Acta Neurobiol Exp (Wars).* 1993;53(1):325-335.
27. Bancher C, Brunner C, Lassmann H, et al. Accumulation of abnormally phosphorylated tau precedes the formation of neurofibrillary tangles in Alzheimer's disease. *Brain Res.* 1989;477(1-2):90-99.
28. Mandelkow EM, Mandelkow E. Tau in Alzheimer's disease. *Trends Cell Biol.* 1998;8(11):425-427.
29. Hampel H, Buerger K, Zinkowski R, et al. Measurement of phosphorylated tau epitopes in the differential diagnosis of Alzheimer disease: a comparative cerebrospinal fluid study. *Arch Gen Psychiatry.* 2004;61(1):95-102.
30. Augustinack JC, Schneider A, Mandelkow EM, Hyman BT. Specific tau phosphorylation sites correlate with severity of neuronal cytopathology in Alzheimer's disease. *Acta Neuropathol.* 2002;103(1):26-35.
31. Winblad B, Amouyel P, Andrieu S, et al. Defeating Alzheimer's disease and other dementias: a priority for European science and society. *Lancet Neurol.* 2016;15(5):455-532.

References

32. Walsh S, Merrick R, Milne R, Brayne C. Aducanumab for Alzheimer's disease? *Bmj*. 2021;374:n1682.
33. Masters CL, Bateman R, Blennow K, Rowe CC, Sperling RA, Cummings JL. Alzheimer's disease. *Nature Reviews Disease Primers*. 2015;1(1):15056.
34. Zagórska A, Jaromin A. Perspectives for New and More Efficient Multifunctional Ligands for Alzheimer's Disease Therapy. *Molecules*. 2020;25(15).
35. Anderson NL, Anderson NG. The Human Plasma Proteome: History, Character, and Diagnostic Prospects*. *Molecular & Cellular Proteomics*. 2002;1(11):845-867.
36. Kadry H, Noorani B, Cucullo L. A blood-brain barrier overview on structure, function, impairment, and biomarkers of integrity. *Fluids Barriers CNS*. 2020;17(1):69.
37. Sweeney MD, Kisler K, Montagne A, Toga AW, Zlokovic BV. The role of brain vasculature in neurodegenerative disorders. *Nat Neurosci*. 2018;21(10):1318-1331.
38. Kaur G, Poljak A, Ali SA, Zhong L, Raftery MJ, Sachdev P. Extending the Depth of Human Plasma Proteome Coverage Using Simple Fractionation Techniques. *Journal of Proteome Research*. 2021;20(2):1261-1279.
39. Aydin S. A short history, principles, and types of ELISA, and our laboratory experience with peptide/protein analyses using ELISA. *Peptides*. 2015;72:4-15.
40. Sakamoto S, Putalun W, Vimolmangkang S, et al. Enzyme-linked immunosorbent assay for the quantitative/qualitative analysis of plant secondary metabolites. *J Nat Med*. 2018;72(1):32-42.
41. Mahmood T, Yang P-C. Western blot: technique, theory, and trouble shooting. *N Am J Med Sci*. 2012;4(9):429-434.
42. Ghosh R, Gilda JE, Gomes AV. The necessity of and strategies for improving confidence in the accuracy of western blots. *Expert Rev Proteomics*. 2014;11(5):549-560.
43. Duraiyan J, Govindarajan R, Kaliyappan K, Palanisamy M. Applications of immunohistochemistry. *J Pharm Bioallied Sci*. 2012;4(Suppl 2):S307-S309.
44. Maher S, Jjunju FPM, Taylor S. Colloquium: 100 years of mass spectrometry: Perspectives and future trends. *Reviews of Modern Physics*. 2015;87(1):113-135.
45. Griffiths J. A Brief History of Mass Spectrometry. *Analytical Chemistry*. 2008;80(15):5678-5683.
46. Lössl P, van de Waterbeemd M, Heck A, Jr. The diverse and expanding role of mass spectrometry in structural and molecular biology. *EMBO J*. 2016;35(24):2634-2657.
47. Siuzdak G. The expanding role of mass spectrometry for biotechnology. 2003.
48. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng*. 2009;11:49-79.
49. Wasinger VC, Cordwell SJ, Cerpa-Poljak A, et al. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis*. 1995;16(7):1090-1094.

References

50. Magdeldin S, Enany S, Yoshida Y, et al. Basics and recent advances of two dimensional-polyacrylamide gel electrophoresis. *Clinical Proteomics*. 2014;11(1):16.
51. Calderaro A, Arcangeletti M-C, Rodighiero I, et al. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry applied to virus identification. *Scientific Reports*. 2014;4(1):6803.
52. Whitehouse CM, Dreyer RN, Yamashita M, Fenn JB. Electrospray interface for liquid chromatographs and mass spectrometers. *Anal Chem*. 1985;57(3):675-679.
53. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003;422(6928):198-207.
54. Liebler DC, Zimmerman LJ. Targeted quantitation of proteins by mass spectrometry. *Biochemistry*. 2013;52(22):3797-3806.
55. Sturm RM, Lietz CB, Li L. Improved isobaric tandem mass tag quantification by ion mobility mass spectrometry. *Rapid Commun Mass Spectrom*. 2014;28(9):1051-1060.
56. Zhu W, Smith JW, Huang C-M. Mass Spectrometry-Based Label-Free Quantitative Proteomics. *Journal of Biomedicine and Biotechnology*. 2010;2010:840518.
57. Ishihama Y, Oda Y, Tabata T, et al. Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein*S. *Molecular & Cellular Proteomics*. 2005;4(9):1265-1272.
58. Fabre B, Lambour T, Bouyssié D, et al. Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry. *EuPA Open Proteomics*. 2014;4:82-86.
59. Sethuraman M, McComb ME, Huang H, et al. Isotope-Coded Affinity Tag (ICAT) Approach to Redox Proteomics: Identification and Quantitation of Oxidant-Sensitive Cysteine Thiols in Complex Protein Mixtures. *Journal of Proteome Research*. 2004;3(6):1228-1233.
60. Thompson A, Schäfer J, Kuhn K, et al. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Analytical Chemistry*. 2003;75(8):1895-1904.
61. Ong SE, Blagoev B, Kratchmarova I, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*. 2002;1(5):376-386.
62. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol*. 2018;14(8):e8126.
63. Cummings J. The National Institute on Aging—Alzheimer's Association Framework on Alzheimer's disease: Application to clinical trials. *Alzheimer's & Dementia*. 2019;15(1):172-178.

References

64. Jiang D, Rauda I, Han S, Chen S, Zhou F. Aggregation Pathways of the Amyloid $\beta(1-42)$ Peptide Depend on Its Colloidal Stability and Ordered β -Sheet Stacking. *Langmuir*. 2012;28(35):12711-12721.
65. Dhiman K, Gupta VB, Villemagne VL, et al. Cerebrospinal fluid neurofilament light concentration predicts brain atrophy and cognition in Alzheimer's disease. *Alzheimers Dement (Amst)*. 2020;12(1):e12005-e12005.
66. Montagne A, Barnes SR, Sweeney MD, et al. Blood-brain barrier breakdown in the aging human hippocampus. *Neuron*. 2015;85(2):296-302.
67. Guadaño-Ferraz A, Viñuela A, Oeding G, Bernal J, Rausell E. RC3/neurogranin is expressed in pyramidal neurons of motor and somatosensory cortex in normal and denervated monkeys. *J Comp Neurol*. 2005;493(4):554-570.
68. Wang L. Association of cerebrospinal fluid Neurogranin with Alzheimer's disease. *Aging Clin Exp Res*. 2019;31(2):185-191.
69. Mavroudis IA, Petridis F, Chatzikonstantinou S, Kazis D. A meta-analysis on CSF neurogranin levels for the diagnosis of Alzheimer's disease and mild cognitive impairment. *Aging Clin Exp Res*. 2020;32(9):1639-1646.
70. Butovsky O, Weiner HL. Microglial signatures and their role in health and disease. *Nat Rev Neurosci*. 2018;19(10):622-635.
71. Kober DL, Brett TJ. TREM2-Ligand Interactions in Health and Disease. *J Mol Biol*. 2017;429(11):1607-1629.
72. Wang Y, Cella M, Mallinson K, et al. TREM2 Lipid Sensing Sustains the Microglial Response in an Alzheimer's Disease Model. *Cell*. 2015;160(6):1061-1071.
73. Hsieh CL, Koike M, Spusta SC, et al. A role for TREM2 ligands in the phagocytosis of apoptotic neuronal cells by microglia. *J Neurochem*. 2009;109(4):1144-1156.
74. Piccio L, Buonsanti C, Cella M, et al. Identification of soluble TREM-2 in the cerebrospinal fluid and its association with multiple sclerosis and CNS inflammation. *Brain*. 2008;131(Pt 11):3081-3091.
75. Kleinberger G, Yamanishi Y, Suárez-Calvet M, et al. TREM2 mutations implicated in neurodegeneration impair cell surface transport and phagocytosis. *Sci Transl Med*. 2014;6(243):243ra286.
76. Schlepckow K, Kleinberger G, Fukumori A, et al. An Alzheimer-associated TREM2 variant occurs at the ADAM cleavage site and affects shedding and phagocytic function. *EMBO Mol Med*. 2017;9(10):1356-1365.
77. Heslegrave A, Heywood W, Paterson R, et al. Increased cerebrospinal fluid soluble TREM2 concentration in Alzheimer's disease. *Mol Neurodegener*. 2016;11:3.

References

78. Piccio L, Deming Y, Del-Águila JL, et al. Cerebrospinal fluid soluble TREM2 is higher in Alzheimer disease and associated with mutation status. *Acta Neuropathol.* 2016;131(6):925-933.
79. Suárez-Calvet M, Kleinberger G, Araque Caballero M, et al. sTREM2 cerebrospinal fluid levels are a potential biomarker for microglia activity in early-stage Alzheimer's disease and associate with neuronal injury markers. *EMBO Mol Med.* 2016;8(5):466-476.
80. Suárez-Calvet M, Morenas-Rodríguez E, Kleinberger G, et al. Early increase of CSF sTREM2 in Alzheimer's disease is associated with tau related-neurodegeneration but not with amyloid- β pathology. *Molecular Neurodegeneration.* 2019;14(1):1.
81. Gispert JD, Suárez-Calvet M, Monté GC, et al. Cerebrospinal fluid sTREM2 levels are associated with gray matter volume increases and reduced diffusivity in early Alzheimer's disease. *Alzheimers Dement.* 2016;12(12):1259-1272.
82. Ewers M, Franzmeier N, Suárez-Calvet M, et al. Increased soluble TREM2 in cerebrospinal fluid is associated with reduced cognitive and clinical decline in Alzheimer's disease. *Sci Transl Med.* 2019;11(507).
83. Ewers M, Biechele G, Suárez-Calvet M, et al. Higher CSF sTREM2 and microglia activation are associated with slower rates of beta-amyloid accumulation. *EMBO Mol Med.* 2020;12(9):e12308.
84. Guerreiro RJ, Lohmann E, Brás JM, et al. Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia-like syndrome without bone involvement. *JAMA Neurol.* 2013;70(1):78-84.
85. Borroni B, Ferrari F, Galimberti D, et al. Heterozygous TREM2 mutations in frontotemporal dementia. *Neurobiology of Aging.* 2014;35(4):934.e937-934.e910.
86. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 Associated with the Risk of Alzheimer's Disease. *New England Journal of Medicine.* 2012;368(2):107-116.
87. Guerreiro R, Wojtas A, Bras J, et al. TREM2 Variants in Alzheimer's Disease. *New England Journal of Medicine.* 2012;368(2):117-127.
88. Roussos P, Katsel P, Fam P, Tan W, Purohit DP, Haroutunian V. The triggering receptor expressed on myeloid cells 2 (TREM2) is associated with enhanced inflammation, neuropathological lesions and increased risk for Alzheimer's dementia. *Alzheimers Dement.* 2015;11(10):1163-1170.
89. Mattsson N, Tabatabaei S, Johansson P, et al. Cerebrospinal Fluid Microglial Markers in Alzheimer's Disease: Elevated Chitotriosidase Activity but Lack of Diagnostic Utility. *NeuroMolecular Medicine.* 2011;13(2):151-159.
90. Bonneh-Barkay D, Bissel SJ, Kofler J, Starkey A, Wang G, Wiley CA. Astrocyte and macrophage regulation of YKL-40 expression and cellular response in neuroinflammation. *Brain Pathol.* 2012;22(4):530-546.

References

91. Craig-Schapiro R, Perrin RJ, Roe CM, et al. YKL-40: A Novel Prognostic Fluid Biomarker for Preclinical Alzheimer's Disease. *Biological Psychiatry*. 2010;68(10):903-912.
92. Sutphen CL, Jasielec MS, Shah AR, et al. Longitudinal Cerebrospinal Fluid Biomarker Changes in Preclinical Alzheimer Disease During Middle Age. *JAMA Neurology*. 2015;72(9):1029-1042.
93. Wang L, Gao T, Cai T, Li K, Zheng P, Liu J. Cerebrospinal fluid levels of YKL-40 in prodromal Alzheimer's disease. *Neuroscience Letters*. 2020;715:134658.
94. Wang L, Gao T, Cai T, Li K, Zheng P, Liu J. Cerebrospinal fluid levels of YKL-40 in prodromal Alzheimer's disease. *Neurosci Lett*. 2020;715:134658.
95. Thordardottir S, Almkvist O, Johansson C, Zetterberg H, Blennow K, Graff C. Cerebrospinal Fluid YKL-40 and Neurogranin in Familial Alzheimer's Disease: A Pilot Study. *Journal of Alzheimer's Disease*. 2020;76:941-953.
96. Tible M, Sandelius Å, Höglund K, et al. Dissection of synaptic pathways through the CSF biomarkers for predicting Alzheimer disease. *Neurology*. 2020;95(8):e953.
97. Ashton NJ, Leuzy A, Lim YM, et al. Increased plasma neurofilament light chain concentration correlates with severity of post-mortem neurofibrillary tangle pathology and neurodegeneration. *Acta Neuropathologica Communications*. 2019;7(1):5.
98. Hampel H, O'Bryant SE, Molinuevo JL, et al. Blood-based biomarkers for Alzheimer disease: mapping the road to the clinic. *Nat Rev Neurol*. 2018;14(11):639-652.
99. Chong JR, Ashton NJ, Karikari TK, et al. Blood-based high sensitivity measurements of beta-amyloid and phosphorylated tau as biomarkers of Alzheimer's disease: a focused review on recent advances. *J Neurol Neurosurg Psychiatry*. 2021;92(11):1231-1241.
100. Kaur G, Poljak A, Braidly N, Crawford JD, Lo J, Sachdev PS. Fluid Biomarkers and APOE Status of Early Onset Alzheimer's Disease Variants: A Systematic Review and Meta-Analysis. *Journal of Alzheimer's Disease*. 2020;75:827-843.
101. Khoury R, Ghossoub E. Diagnostic biomarkers of Alzheimer's disease: A state-of-the-art review. *Biomarkers in Neuropsychiatry*. 2019;1:100005.
102. Zetterberg H, Schott JM. Biomarkers for Alzheimer's disease beyond amyloid and tau. *Nature Medicine*. 2019;25(2):201-203.
103. Hampel H, O'Bryant SE, Molinuevo JL, et al. Blood-based biomarkers for Alzheimer disease: mapping the road to the clinic. *Nature Reviews Neurology*. 2018;14(11):639-652.
104. Vassar R, Bennett BD, Babu-Khan S, et al. β -Secretase Cleavage of Alzheimer's Amyloid Precursor Protein by the Transmembrane Aspartic Protease BACE. *Science*. 1999;286(5440):735-741.
105. Cervellati C, Trentini A, Rosta V, et al. Serum beta-secretase 1 (BACE1) activity as candidate biomarker for late-onset Alzheimer's disease. *Geroscience*. 2020;42(1):159-167.

References

106. Shen Y, Wang H, Sun Q, et al. Increased Plasma Beta-Secretase 1 May Predict Conversion to Alzheimer's Disease Dementia in Individuals With Mild Cognitive Impairment. *Biological Psychiatry*. 2018;83(5):447-455.
107. Wu G, Sankaranarayanan S, Wong J, et al. Characterization of plasma β -secretase (BACE1) activity and soluble amyloid precursor proteins as potential biomarkers for Alzheimer's disease. *J Neurosci Res*. 2012;90(12):2247-2258.
108. Villemagne VL, Fodero-Tavoletti MT, Masters CL, Rowe CC. Tau imaging: early progress and future directions. *The Lancet Neurology*. 2015;14(1):114-124.
109. Filippi L, Chiaravalloti A, Bagni O, Schillaci O. (18)F-labeled radiopharmaceuticals for the molecular neuroimaging of amyloid plaques in Alzheimer's disease. *Am J Nucl Med Mol Imaging*. 2018;8(4):268-281.
110. Suppiah S, Didier MA, Vinjamuri S. The Who, When, Why, and How of PET Amyloid Imaging in Management of Alzheimer's Disease-Review of Literature and Interesting Images. *Diagnostics (Basel)*. 2019;9(2).
111. Hanseeuw BJ, Betensky RA, Jacobs HIL, et al. Association of Amyloid and Tau With Cognition in Preclinical Alzheimer Disease: A Longitudinal Study. *JAMA Neurology*. 2019;76(8):915-924.
112. Schwarz AJ, Yu P, Miller BB, et al. Regional profiles of the candidate tau PET ligand 18 F-AV-1451 recapitulate key features of Braak histopathological stages. *Brain*. 2016;139(5):1539-1550.
113. Yap SY, Frias B, Wren MC, et al. Discriminatory ability of next-generation tau PET tracers for Alzheimer's disease. *Brain*. 2021;144(8):2284-2290.
114. Wilson H, Pagano G, Politis M. Dementia spectrum disorders: lessons learnt from decades with PET research. *Journal of Neural Transmission*. 2019;126(3):233-251.
115. Werry EL, Bright FM, Piguet O, et al. Recent Developments in TSPO PET Imaging as A Biomarker of Neuroinflammation in Neurodegenerative Disorders. *International Journal of Molecular Sciences*. 2019;20(13):3161.
116. Márquez F, Yassa MA. Neuroimaging Biomarkers for Alzheimer's Disease. *Molecular Neurodegeneration*. 2019;14(1):21.
117. Morris GP, Clark IA, Vissel B. Inconsistencies and controversies surrounding the amyloid hypothesis of Alzheimer's disease. *Acta neuropathologica communications*. 2014;2:135-135.
118. Gleeup HS, Hasselbalch SG, Simonsen AH. Biomarkers for Alzheimer's Disease in Saliva: A Systematic Review. *Disease markers*. 2019;2019:4761054-4761054.
119. Lee M, Guo J-P, Kennedy K, McGeer EG, McGeer PL. A Method for Diagnosing Alzheimer's Disease Based on Salivary Amyloid- β Protein 42 Levels. *Journal of Alzheimer's Disease*. 2017;55:1175-1182.

References

120. Sabbagh MN, Shi J, Lee M, et al. Salivary beta amyloid protein levels are detectable and differentiate patients with Alzheimer's disease dementia from normal controls: preliminary findings. *BMC Neurol.* 2018;18(1):155-155.
121. Bermejo-Pareja F, Antequera D, Vargas T, Molina JA, Carro E. Saliva levels of Abeta1-42 as potential biomarker of Alzheimer's disease: a pilot study. *BMC Neurol.* 2010;10(1):108.
122. McGeer PL, Guo JP, Lee M, Kennedy K, McGeer EG. Alzheimer's Disease Can Be Spared by Nonsteroidal Anti-Inflammatory Drugs. *Journal of Alzheimer's Disease.* 2018;62:1219-1222.
123. Ashton NJ, Ide M, Zetterberg H, Blennow K. Salivary Biomarkers for Alzheimer's Disease and Related Disorders. *Neurol Ther.* 2019;8(Suppl 2):83-94.
124. Shi M, Sui Y-T, Peskind ER, et al. Salivary Tau Species are Potential Biomarkers of Alzheimer's Disease. *Journal of Alzheimer's Disease.* 2011;27:299-305.
125. Lau H-C, Lee I-K, Ko P-W, et al. Non-Invasive Screening for Alzheimer's Disease by Sensing Salivary Sugar Using Drosophila Cells Expressing Gustatory Receptor (Gr5a) Immobilized on an Extended Gate Ion-Sensitive Field-Effect Transistor (EG-ISFET) Biosensor. *PLOS ONE.* 2015;10(2):e0117810.
126. Lewczuk P, Riederer P, O'Bryant SE, et al. Cerebrospinal fluid and blood biomarkers for neurodegenerative dementias: An update of the Consensus of the Task Force on Biological Markers in Psychiatry of the World Federation of Societies of Biological Psychiatry. *World J Biol Psychiatry.* 2018;19(4):244-328.
127. Rehiman SH, Lim SM, Neoh CF, et al. Proteomics as a reliable approach for discovery of blood-based Alzheimer's disease biomarkers: A systematic review and meta-analysis. *Ageing Research Reviews.* 2020;60:101066.
128. Eikelenboom P, Hack CE, Rozemuller JM, Stam FC. Complement activation in amyloid plaques in Alzheimer's dementia. *Virchows Archiv B Cell Pathology Including Molecular Pathology.* 1988;56(1):259-262.
129. Hall JR, Wiechmann AR, Johnson LA, Edwards ML, O'Bryant SE. Levels of α -2 macroglobulin in cognitively normal Mexican-Americans with subjective cognitive decline: a HABLE study. *Curr Neurobiol.* 2019;10:22-25.
130. Saunders AJ, Bertram L, Mullin K, et al. Genetic association of Alzheimer's disease with multiple polymorphisms in alpha-2-macroglobulin. *Human Molecular Genetics.* 2003;12(21):2765-2776.
131. Blacker D, Wilcox MA, Laird NM, et al. Alpha-2 macroglobulin is genetically associated with Alzheimer disease. *Nature Genetics.* 1998;19(4):357-360.
132. Kratzer I, Bernhart E, Wintersperger A, et al. Afamin is synthesized by cerebrovascular endothelial cells and mediates α -tocopherol transport across an in vitro model of the blood-brain barrier. *Journal of Neurochemistry.* 2009;108(3):707-718.

References

133. Kronenberg F, Dieplinger H. Afamin is a promising novel marker for metabolic syndrome and related diseases. *Clinical Lipidology*. 2015;10(3):207-210.
134. Kim J, Basak JM, Holtzman DM. The Role of Apolipoprotein E in Alzheimer's Disease. *Neuron*. 2009;63(3):287-303.
135. Sengupta MB, Saha S, Mohanty PK, Mukhopadhyay KK, Mukhopadhyay D. Increased expression of ApoA1 after neuronal injury may be beneficial for healing. *Molecular and Cellular Biochemistry*. 2017;424(1-2):45-55.
136. Khan D, Vasu S, Moffett RC, Irwin N, Flatt PR. Influence of neuropeptide Y and pancreatic polypeptide on islet function and beta-cell survival. *Biochimica et Biophysica Acta - General Subjects*. 2017;1861(4):749-758.
137. Kouli A, Camacho M, Allinson K, Williams-Gray CH. Neuroinflammation and protein pathology in Parkinson's disease dementia. *Acta Neuropathologica Communications*. 2020;8(1):211.
138. Rabilloud T. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics*. 2002;2(1):3-10.
139. Patton WF. Detection technologies in proteome analysis. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2002;771(1-2):3-31.
140. Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP, Veenstra TD. Characterization of the Low Molecular Weight Human Serum Proteome*S. *Molecular & Cellular Proteomics*. 2003;2(10):1096-1103.
141. Chandramouli K, Qian P-Y. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Hum Genomics Proteomics*. 2009;2009:239204.
142. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):280-292.
143. Sperling RA, Jack CR, Jr., Aisen PS. Testing the right target and right drug at the right stage. *Sci Transl Med*. 2011;3(111):111cm133.
144. Bai B, Wang X, Li Y, et al. Deep Multilayer Brain Proteomics Identifies Molecular Networks in Alzheimer's Disease Progression. *Neuron*. 2020;105(6):975-991.e977.
145. Isik AT. Late onset Alzheimer's disease in older people. *Clin Interv Aging*. 2010;5:307-311.
146. Bateman RJ, Aisen PS, De Strooper B, et al. Autosomal-dominant Alzheimer's disease: a review and proposal for the prevention of Alzheimer's disease. *Alzheimers Res Ther*. 2011;3(1):1.
147. Carmona S, Hardy J, Guerreiro R. The genetic landscape of Alzheimer disease. *Handb Clin Neurol*. 2018;148:395-408.

References

148. Holtzman DM, Herz J, Bu G. Apolipoprotein E and apolipoprotein E receptors: normal biology and roles in Alzheimer disease. *Cold Spring Harb Perspect Med.* 2012;2(3):a006312.
149. Safieh M, Korczyn AD, Michaelson DM. ApoE4: an emerging therapeutic target for Alzheimer's disease. *BMC Medicine.* 2019;17(1):64.
150. Kaur G, Poljak A, Braidy N, Crawford JD, Lo J, Sachdev PS. Fluid Biomarkers and APOE Status of Early Onset Alzheimer's Disease Variants: A Systematic Review and Meta-Analysis. *J Alzheimers Dis.* 2020;75(3):827-843.
151. de Chaves EP, Narayanaswami V. Apolipoprotein E and cholesterol in aging and disease in the brain. *Future Lipidol.* 2008;3(5):505-530.
152. Bu G. Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy. *Nat Rev Neurosci.* 2009;10(5):333-344.
153. Rabinovici GD. Late-onset Alzheimer Disease. *Continuum (Minneap Minn).* 2019;25(1):14-33.
154. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology.* 1999;56(3):303-308.
155. Higginbotham L, Ping L, Dammer EB, et al. Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic Alzheimer's disease. *Sci Adv.* 2020;6(43).
156. Games D, Adams D, Alessandrini R, et al. Alzheimer-type neuropathology in transgenic mice overexpressing V717F beta-amyloid precursor protein. *Nature.* 1995;373(6514):523-527.
157. Herholz K. Cerebral glucose metabolism in preclinical and prodromal Alzheimer's disease. *Expert Rev Neurother.* 2010;10(11):1667-1673.
158. van der Velpen V, Teav T, Gallart-Ayala H, et al. Systemic and central nervous system metabolic alterations in Alzheimer's disease. *Alzheimer's Research & Therapy.* 2019;11(1):93.
159. Martin LJ. Mitochondrial and Cell Death Mechanisms in Neurodegenerative Diseases. *Pharmaceuticals (Basel).* 2010;3(4):839-915.
160. Peña-Oyarzun D, Bravo-Sagua R, Diaz-Vega A, et al. Autophagy and oxidative stress in non-communicable diseases: A matter of the inflammatory state? *Free Radic Biol Med.* 2018;124:61-78.
161. Liu J, Li L. Targeting Autophagy for the Treatment of Alzheimer's Disease: Challenges and Opportunities. *Frontiers in molecular neuroscience.* 2019;12:203-203.
162. Boland B, Kumar A, Lee S, et al. Autophagy induction and autophagosome clearance in neurons: relationship to autophagic pathology in Alzheimer's disease. *J Neurosci.* 2008;28(27):6926-6937.

References

163. Ferreira ST, Clarke JR, Bomfim TR, De Felice FG. Inflammation, defective insulin signaling, and neuronal dysfunction in Alzheimer's disease. *Alzheimers Dement.* 2014;10(1 Suppl):S76-83.
164. Herrero MT, Estrada C, Maatouk L, Vyas S. Inflammation in Parkinson's disease: role of glucocorticoids. *Front Neuroanat.* 2015;9:32.
165. Breunig J, Guillot-Sestier M-V, Town T. Brain injury, neuroinflammation and Alzheimer's disease. *Front Aging Neurosci.* 2013;5(26).
166. McCombe PA, Henderson RD. The Role of immune and inflammatory mechanisms in ALS. *Curr Mol Med.* 2011;11(3):246-254.
167. Heneka MT, Golenbock DT, Latz E. Innate immunity in Alzheimer's disease. *Nat Immunol.* 2015;16(3):229-236.
168. Huang W-J, Zhang X, Chen W-W. Role of oxidative stress in Alzheimer's disease. *Biomed Rep.* 2016;4(5):519-522.
169. Cheng J, Wang S, Dong Y, Yuan Z. The Role and Regulatory Mechanism of Hippo Signaling Components in the Neuronal System. *Frontiers in Immunology.* 2020;11:281.
170. Foote M, Zhou Y. 14-3-3 proteins in neurological disorders. *Int J Biochem Mol Biol.* 2012;3(2):152-164.
171. Higginbotham L, Ping L, Dammer EB, et al. Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic Alzheimer's disease. *Science Advances.* 2020;6(43):eaaz9360.
172. Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol.* 2009;65(4):403-413.
173. Wesseling H, Mair W, Kumar M, et al. Tau PTM Profiles Identify Patient Heterogeneity and Stages of Alzheimer's Disease. *Cell.* 2020;183(6):1699-1713.e1613.
174. Min SW, Cho SH, Zhou Y, et al. Acetylation of tau inhibits its degradation and contributes to tauopathy. *Neuron.* 2010;67(6):953-966.
175. Cohen TJ, Guo JL, Hurtado DE, et al. The acetylation of tau inhibits its function and promotes pathological tau aggregation. *Nature Communications.* 2011;2(1):252.
176. Abreha MH, Dammer EB, Ping L, et al. Quantitative Analysis of the Brain Ubiquitylome in Alzheimer's Disease. *Proteomics.* 2018;18(20):e1800108.
177. Dammer EB, Na CH, Xu P, et al. Polyubiquitin linkage profiles in three models of proteolytic stress suggest the etiology of Alzheimer disease. *J Biol Chem.* 2011;286(12):10457-10465.
178. Dai MH, Zheng H, Zeng LD, Zhang Y. The genes associated with early-onset Alzheimer's disease. *Oncotarget.* 2018;9(19):15132-15143.

References

179. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700.
180. McDade E, Wang G, Gordon BA, et al. Longitudinal cognitive and biomarker changes in dominantly inherited Alzheimer disease. *Neurology*. 2018;91(14):e1295-e1306.
181. Fleisher AS, Chen K, Quiroz YT, et al. Associations between biomarkers and age in the presenilin 1 E280A autosomal dominant Alzheimer disease kindred: a cross-sectional study. *JAMA Neurol*. 2015;72(3):316-324.
182. Lim YY, Hassenstab J, Cruchaga C, et al. BDNF Val66Met moderates memory impairment, hippocampal function and tau in preclinical autosomal dominant Alzheimer's disease. *Brain*. 2016;139(Pt 10):2766-2777.
183. Serghiou S, Goodman SN. Random-Effects Meta-analysis: Summarizing Evidence With Caveats. *Jama*. 2019;321(3):301-302.
184. Sedgwick P. Meta-analyses: how to read a funnel plot. *BMJ : British Medical Journal*. 2013;346:f1342.
185. Schoonenboom NS, Pijnenburg YA, Mulder C, et al. Amyloid beta(1-42) and phosphorylated tau in CSF as markers for early-onset Alzheimer disease. *Neurology*. 2004;62(9):1580-1584.
186. de Jong D, Jansen RW, Pijnenburg YA, et al. CSF neurofilament proteins in the differential diagnosis of dementia. *J Neurol Neurosurg Psychiatry*. 2007;78(9):936-938.
187. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology*. 2005;5(1):13.
188. Schöll M, Wall A, Thordardottir S, et al. Low PiB PET retention in presence of pathologic CSF biomarkers in Arctic APP mutation carriers. *Neurology*. 2012;79(3):229-236.
189. Preische O, Schultz SA, Apel A, et al. Serum neurofilament dynamics predicts neurodegeneration and clinical progression in presymptomatic Alzheimer's disease. *Nat Med*. 2019;25(2):277-283.
190. Weston PSJ, Poole T, Ryan NS, et al. Serum neurofilament light in familial Alzheimer disease: A marker of early neurodegeneration. *Neurology*. 2017;89(21):2167-2175.
191. Ikeda M, Yonemura K, Kakuda S, et al. Cerebrospinal fluid levels of phosphorylated tau and A β 1-38/A β 1-40/A β 1-42 in Alzheimer's disease with PS1 mutations. *Amyloid*. 2013;20(2):107-112.
192. Dorey A, Perret-Liaudet A, Tholance Y, Fourier A, Quadrio I. Cerebrospinal Fluid A β 40 Improves the Interpretation of A β 42 Concentration for Diagnosing Alzheimer's Disease. *Front Neurol*. 2015;6:247-247.
193. Lashley T, Schott JM, Weston P, et al. Molecular biomarkers of Alzheimer's disease: progress and prospects. *Dis Model Mech*. 2018;11(5).

References

194. Yuan A, Sershen H, Veeranna, et al. Neurofilament subunits are integral components of synapses and modulate neurotransmission and behavior in vivo. *Mol Psychiatry*. 2015;20(8):986-994.
195. Bacioglu M, Maia LF, Preische O, et al. Neurofilament Light Chain in Blood and CSF as Marker of Disease Progression in Mouse Models and in Neurodegenerative Diseases. *Neuron*. 2016;91(1):56-66.
196. Zetterberg H. Neurofilament Light: A Dynamic Cross-Disease Fluid Biomarker for Neurodegeneration. *Neuron*. 2016;91(1):1-3.
197. Kester MI, Teunissen CE, Sutphen C, et al. Cerebrospinal fluid VILIP-1 and YKL-40, candidate biomarkers to diagnose, predict and monitor Alzheimer's disease in a memory clinic cohort. *Alzheimers Res Ther*. 2015;7(1):59.
198. Song F, Poljak A, Valenzuela M, Mayeux R, Smythe GA, Sachdev PS. Meta-analysis of plasma amyloid- β levels in Alzheimer's disease. *J Alzheimers Dis*. 2011;26(2):365-375.
199. Herukka SK, Simonsen AH, Andreasen N, et al. Recommendations for cerebrospinal fluid Alzheimer's disease biomarkers in the diagnostic evaluation of mild cognitive impairment. *Alzheimers Dement*. 2017;13(3):285-295.
200. Thambisetty M, Lovestone S. Blood-based biomarkers of Alzheimer's disease: challenging but feasible. *Biomark Med*. 2010;4(1):65-79.
201. Mayeux R, Honig LS, Tang MX, et al. Plasma A[β]40 and A[β]42 and Alzheimer's disease: relation to age, mortality, and risk. *Neurology*. 2003;61(9):1185-1190.
202. Findeis MA. The role of amyloid beta peptide 42 in Alzheimer's disease. *Pharmacol Ther*. 2007;116(2):266-286.
203. Lundström SL, Yang H, Lyutvinskiy Y, et al. Blood plasma IgG Fc glycans are significantly altered in Alzheimer's disease and progressive mild cognitive impairment. *J Alzheimers Dis*. 2014;38(3):567-579.
204. Muenchhoff J, Poljak A, Thalamuthu A, et al. Changes in the plasma proteome at asymptomatic and symptomatic stages of autosomal dominant Alzheimer's disease. *Sci Rep*. 2016;6:29078.
205. Hong S, Beja-Glasser VF, Nfonoyim BM, et al. Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science*. 2016;352(6286):712-716.
206. Chung WS, Verghese PB, Chakraborty C, et al. Novel allele-dependent role for APOE in controlling the rate of synapse pruning by astrocytes. *Proc Natl Acad Sci U S A*. 2016;113(36):10186-10191.
207. Huynh RA, Mohan C. Alzheimer's Disease: Biomarkers in the Genome, Blood, and Cerebrospinal Fluid. *Front Neurol*. 2017;8:102.
208. Vemuri P, Wiste HJ, Weigand SD, et al. Effect of apolipoprotein E on biomarkers of amyloid load and neuronal pathology in Alzheimer disease. *Ann Neurol*. 2010;67(3):308-316.

References

209. Végvári A, Welinder C, Lindberg H, Fehniger TE, Marko-Varga G. Biobank resources for future patient care: developments, principles and concepts. *J Clin Bioinforma*. 2011;1(1):24.
210. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics*. 2002;1(11):845-867.
211. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol*. 2006;24(8):971-983.
212. Geyer PE, Holdt LM, Teupser D, Mann M. Revisiting biomarker discovery by plasma proteomics. *Mol Syst Biol*. 2017;13(9):942.
213. Anderson L. Six decades searching for meaning in the proteome. *J Proteomics*. 2014;107:24-30.
214. Omenn GS. Exploring the human plasma proteome. *Proteomics*. 2005;5(13):3223, 3225.
215. Cox J, Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem*. 2011;80:273-299.
216. Geiger T, Cox J, Mann M. Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol Cell Proteomics*. 2010;9(10):2252-2261.
217. Munoz J, Heck AJ. From the human genome to the human proteome. *Angewandte Chemie International Edition*. 2014;53(41):10864-10866.
218. Sturgeon C, Hill R, Hortin GL, Thompson D. Taking a new biomarker into routine use--a perspective from the routine clinical biochemistry laboratory. *Proteomics Clin Appl*. 2010;4(12):892-903.
219. Vihko P, Kontturi M, Korhonen LK. Purification of human prostatic acid phosphatase by affinity chromatography and isoelectric focusing. Part I. *Clin Chem*. 1978;24(3):466-470.
220. Pieper R, Su Q, Gatlin CL, Huang ST, Anderson NL, Steiner S. Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics*. 2003;3(4):422-432.
221. Qian WJ, Kaleta DT, Petritis BO, et al. Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy. *Mol Cell Proteomics*. 2008;7(10):1963-1973.
222. Tu C, Rudnick PA, Martinez MY, et al. Depletion of abundant plasma proteins and limitations of plasma proteomics. *Journal of proteome research*. 2010;9(10):4982-4991.
223. Palstrøm NB, Rasmussen LM, Beck HC. Affinity Capture Enrichment versus Affinity Depletion: A Comparison of Strategies for Increasing Coverage of Low-Abundant Human Plasma Proteins. *International Journal of Molecular Sciences*. 2020;21(16):5903.
224. Babačić H, Lehtiö J, Pico de Coaña Y, Pernemalm M, Eriksson H. In-depth plasma proteomics reveals increase in circulating PD-1 during anti-PD-1 immunotherapy in patients with metastatic cutaneous melanoma. *J Immunother Cancer*. 2020;8(1).

References

225. Garcia-Rosa S, Carvalho BS, Guest PC, Steiner J, Martins-de-Souza D. Blood plasma proteomic modulation induced by olanzapine and risperidone in schizophrenia patients. *J Proteomics*. 2020;224:103813.
226. Xu R, Gong CX, Duan CM, et al. Age-Dependent Changes in the Plasma Proteome of Healthy Adults. *J Nutr Health Aging*. 2020;24(8):846-856.
227. Zhang Y, Mao Y, Zhao W, et al. Glyco-CPLL: An Integrated Method for In-Depth and Comprehensive N-Glycoproteome Profiling of Human Plasma. *Journal of Proteome Research*. 2020;19(2):655-666.
228. Henriksson AE, Lindqvist M, Sihlbom C, Bergström J, Bylund D. Identification of potential plasma biomarkers for abdominal aortic aneurysm using tandem mass tag quantitative proteomics. *Proteomes*. 2018;6(4):43.
229. Dayon L, Cominetti O, Wojcik Jrm, et al. Proteomes of paired human cerebrospinal fluid and plasma: relation to blood–brain barrier permeability in older adults. *Journal of proteome research*. 2019;18(3):1162-1174.
230. Manfredi M, Chiariello C, Conte E, et al. Plasma Proteome Profiles of Stable CAD Patients Stratified According to Total Apo C-III Levels. *PROTEOMICS–Clinical Applications*. 2019;13(3):1800023.
231. Liu CW, Bramer L, Webb-Robertson BJ, Waugh K, Rewers MJ, Zhang Q. Temporal expression profiling of plasma proteins reveals oxidative stress in early stages of Type 1 Diabetes progression. *J Proteomics*. 2018;172:100-110.
232. Zhou N, Wang K, Fang S, et al. Discovery of a Potential Plasma Protein Biomarker Panel for Acute-on-Chronic Liver Failure Induced by Hepatitis B Virus. *Front Physiol*. 2017;8:1009.
233. Geyer PE, Wewer Albrechtsen NJ, Tyanova S, et al. Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Molecular systems biology*. 2016;12(12):901.
234. Albrechtsen NJW, Geyer PE, Doll S, et al. Plasma proteome profiling reveals dynamics of inflammatory and lipid homeostasis markers after Roux-en-Y gastric bypass surgery. *Cell systems*. 2018;7(6):601-612. e603.
235. Keshishian H, Burgess MW, Gillette MA, et al. Multiplexed, quantitative workflow for sensitive biomarker discovery in plasma yields novel candidates for early myocardial injury. *Molecular & Cellular Proteomics*. 2015;14(9):2375-2393.
236. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature protocols*. 2016;11(12):2301-2319.
237. Prikas E, Poljak A, Ittner A. Mapping p38 α mitogen-activated protein kinase signaling by proximity-dependent labeling. *Protein Science*. 2020;29(5):1196-1210.

References

238. Kim S, Kim D, Cho SW, Kim J, Kim JS. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 2014;24(6):1012-1019.
239. Schwenk JM, Omenn GS, Sun Z, et al. The human plasma proteome draft of 2017: building on the human plasma PeptideAtlas from mass spectrometry and complementary assays. *Journal of proteome research.* 2017;16(12):4299-4310.
240. Li N, Zhou Y, Wang J, et al. Sequential precipitation and delipidation enables efficient enrichment of low-molecular weight proteins and peptides from human plasma. *Journal of Proteome Research.* 2020;19(8):3340-3351.
241. Farrah T, Deutsch EW, Omenn GS, et al. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Molecular & cellular proteomics.* 2011;10(9).
242. Jankovska E, Svitek M, Holada K, Petrak J. Affinity depletion versus relative protein enrichment: a side-by-side comparison of two major strategies for increasing human cerebrospinal fluid proteome coverage. *Clinical proteomics.* 2019;16(1):1-10.
243. Chen L, Zhao W, He J, et al. Label-Free Quantitative Proteomic Screening of Candidate Plasma Biomarkers for the Prognosis of Breast Cancer with Different Lymph Node Statuses. *PROTEOMICS–Clinical Applications.* 2018;12(3):1700117.
244. Choksawangkarn W, Edwards N, Wang Y, Gutierrez P, Fenselau C. Comparative study of workflows optimized for in-gel, in-solution, and on-filter proteolysis in the analysis of plasma membrane proteins. *Journal of proteome research.* 2012;11(5):3030-3034.
245. Geiger T, Wehner A, Schaab C, Cox J, Mann M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Molecular & Cellular Proteomics.* 2012;11(3).
246. Li S, Sheng J, Hu JK, et al. Ribonuclease 4 protects neuron degeneration by promoting angiogenesis, neurogenesis, and neuronal survival under stress. *Angiogenesis.* 2013;16(2):387-404.
247. Wang L, Zhang ZG, Zhang RL, et al. Matrix metalloproteinase 2 (MMP2) and MMP9 secreted by erythropoietin-activated endothelial cells promote neural progenitor cell migration. *Journal of Neuroscience.* 2006;26(22):5996-6003.
248. Ignjatovic V, Geyer PE, Palaniappan KK, et al. Mass spectrometry-based plasma proteomics: considerations from sample collection to achieving translational data. *Journal of proteome research.* 2019;18(12):4085-4097.
249. Surinova S, Schiess R, Hüttenhain R, Cerciello F, Wollscheid B, Aebersold R. On the Development of Plasma Protein Biomarkers. *Journal of Proteome Research.* 2011;10(1):5-16.

References

250. Selkoe DJ. Alzheimer's disease: genes, proteins, and therapy. *Physiol Rev.* 2001;81(2):741-766.
251. Reddy PH, Manczak M, Mao P, Calkins MJ, Reddy AP, Shirendeb U. Amyloid-beta and mitochondria in aging and Alzheimer's disease: implications for synaptic damage and cognitive decline. *J Alzheimers Dis.* 2010;20 Suppl 2(Suppl 2):S499-512.
252. Swerdlow RH, Khan SM. The Alzheimer's disease mitochondrial cascade hypothesis: An update. *Experimental Neurology.* 2009;218(2):308-315.
253. Du H, Guo L, Yan S, Sosunov AA, McKhann GM, ShiDu Yan S. Early deficits in synaptic mitochondria in an Alzheimer's disease mouse model. *Proceedings of the National Academy of Sciences.* 2010;107(43):18670.
254. Xia X, Jiang Q, McDermott J, Han J-DJ. Aging and Alzheimer's disease: Comparison and associations from molecular to system level. *Aging cell.* 2018;17(5):e12802-e12802.
255. Eldholm RS, Persson K, Barca ML, et al. Association between vascular comorbidity and progression of Alzheimer's disease: a two-year observational study in Norwegian memory clinics. *BMC Geriatr.* 2018;18(1):120-120.
256. Mortimer JA, French LR, Hutton JT, Schuman LM. Head injury as a risk factor for Alzheimer's disease. *Neurology.* 1985;35(2):264-267.
257. Fleminger S, Oliver DL, Lovestone S, Rabe-Hesketh S, Giora A. Head injury as a risk factor for Alzheimer's disease: the evidence 10 years on; a partial replication. *J Neurol Neurosurg Psychiatry.* 2003;74(7):857-862.
258. Michaud TL, Su D, Siahpush M, Murman DL. The Risk of Incident Mild Cognitive Impairment and Progression to Dementia Considering Mild Cognitive Impairment Subtypes. *Dement Geriatr Cogn Dis Extra.* 2017;7(1):15-29.
259. Roberts RO, Knopman DS, Mielke MM, et al. Higher risk of progression to dementia in mild cognitive impairment cases who revert to normal. *Neurology.* 2014;82(4):317-325.
260. Keshishian H, Burgess MW, Gillette MA, et al. Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury*. *Molecular & Cellular Proteomics.* 2015;14(9):2375-2393.
261. Sachdev PS, Brodaty H, Reppermund S, et al. The Sydney Memory and Ageing Study (MAS): methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70-90 years. *Int Psychogeriatr.* 2010;22(8):1248-1264.
262. Petersen RC, Doody R, Kurz A, et al. Current concepts in mild cognitive impairment. *Arch Neurol.* 2001;58(12):1985-1992.
263. Csukly G, Sirály E, Fodor Z, et al. The Differentiation of Amnesic Type MCI from the Non-Amnesic Types by Structural MRI. *Front Aging Neurosci.* 2016;8(52).

References

264. Aitken A. Proteins and Proteomics: A Laboratory Manual. R. J. SIMPSON. Cold Spring Harbor Laboratory Press. 2003. 926 pages. ISBN 0 87969 554 4. Price £129. (paperback). ISBN 0 87969 553 6. Price £185. (hardback). *Genetical Research*. 2003;81(3):243-244.
265. Ramus C, Hovasse A, Marcellin M, et al. Benchmarking quantitative label-free LC-MS data processing workflows using a complex spiked proteomic standard dataset. *J Proteomics*. 2016;132:51-62.
266. Ramus C, Hovasse A, Marcellin M, et al. Spiked proteomic standard dataset for testing label-free quantitative software and statistical methods. *Data in Brief*. 2016;6:286-294.
267. Oliveros JC. VENNY. An interactive tool for comparing lists with Venn diagrams. <http://bioinfogpcnbcscics/tools/venny/index.html>. 2007.
268. Wyss-Coray T. Ageing, neurodegeneration and brain rejuvenation. *Nature*. 2016;539(7628):180-186.
269. Elster D, von Eyss B. Hippo signaling in regeneration and aging. *Mechanisms of Ageing and Development*. 2020;189:111280.
270. Saez I, Vilchez D. The Mechanistic Links Between Proteasome Activity, Aging and Age-related Diseases. *Curr Genomics*. 2014;15(1):38-51.
271. Bader JM, Geyer PE, Müller JB, et al. Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease. *Mol Syst Biol*. 2020;16(6):e9356.
272. Cristóvão JS, Gomes CM. S100 Proteins in Alzheimer's Disease. *Frontiers in Neuroscience*. 2019;13(463).
273. Wang T. Glutathione S-transferases variants as risk factors in Alzheimer's disease. *Neurol Sci*. 2015;36(10):1785-1792.
274. Yuan Y, Cruzat VF, Newsholme P, Cheng J, Chen Y, Lu Y. Regulation of SIRT1 in aging: Roles in mitochondrial function and biogenesis. *Mech Ageing Dev*. 2016;155:10-21.
275. Leiser SF, Kaeberlein M. The hypoxia-inducible factor HIF-1 functions as both a positive and negative modulator of aging. *Biol Chem*. 2010;391(10):1131-1137.
276. Moaddel R, Ubaida-Mohien C, Tanaka T, et al. Proteomics in aging research: A roadmap to clinical, translational research. *Aging Cell*. 2021;20(4):e13325.
277. Hoyer S, Oesterreich K, Wagner O. Glucose metabolism as the site of the primary abnormality in early-onset dementia of Alzheimer type? *Journal of Neurology*. 1988;235(3):143-148.
278. Kalaria RN, Harik SI. Reduced glucose transporter at the blood-brain barrier and in cerebral cortex in Alzheimer disease. *J Neurochem*. 1989;53(4):1083-1088.
279. Reiman EM, Caselli RJ, Yun LS, et al. Preclinical Evidence of Alzheimer's Disease in Persons Homozygous for the $\epsilon 4$ Allele for Apolipoprotein E. *New England Journal of Medicine*. 1996;334(12):752-758.
280. Dassati S, Waldner A, Schweigreiter R. Apolipoprotein D takes center stage in the stress response of the aging and degenerative brain. *Neurobiol Aging*. 2014;35(7):1632-1642.

References

281. Rassart E, Bedirian A, Do Carmo S, et al. Apolipoprotein D. *Biochim Biophys Acta*. 2000;1482(1-2):185-198.
282. Perrotte M, Le Page A, Fournet M, et al. Blood-based redox-signature and their association to the cognitive scores in MCI and Alzheimer's disease patients. *Free Radic Biol Med*. 2019;130:499-511.
283. Majima T, Ogita H, Yamada T, et al. Involvement of afadin in the formation and remodeling of synapses in the hippocampus. *Biochemical and Biophysical Research Communications*. 2009;385(4):539-544.
284. Ly CV, Verstreken P. Mitochondria at the synapse. *Neuroscientist*. 2006;12(4):291-299.
285. Nikhil K, Viccaro K, Shah K. Multifaceted Regulation of ALDH1A1 by Cdk5 in Alzheimer's Disease Pathogenesis. *Molecular neurobiology*. 2019;56(2):1366-1390.
286. Styr B, Slutsky I. Imbalance between firing homeostasis and synaptic plasticity drives early-phase Alzheimer's disease. *Nat Neurosci*. 2018;21(4):463-473.
287. Frere S, Slutsky I. Alzheimer's Disease: From Firing Instability to Homeostasis Network Collapse. *Neuron*. 2018;97(1):32-58.
288. Connor JR, Snyder BS, Beard JL, Fine RE, Mufson EJ. Regional distribution of iron and iron-regulatory proteins in the brain in aging and Alzheimer's disease. *J Neurosci Res*. 1992;31(2):327-335.
289. Chen M, Xia W. Proteomic Profiling of Plasma and Brain Tissue from Alzheimer's Disease Patients Reveals Candidate Network of Plasma Biomarkers. *Journal of Alzheimer's disease : JAD*. 2020;76(1):349-368.
290. Ostrowski M, Carmo NB, Krumeich S, et al. Rab27a and Rab27b control different steps of the exosome secretion pathway. *Nature Cell Biology*. 2010;12(1):19-30.
291. Ghidoni R, Paterlini A, Albertini V, et al. Cystatin C is released in association with exosomes: A new tool of neuronal communication which is unbalanced in Alzheimer's disease. *Neurobiology of Aging*. 2011;32(8):1435-1442.
292. Ginsberg SD, Mufson EJ, Alldred MJ, et al. Upregulation of select rab GTPases in cholinergic basal forebrain neurons in mild cognitive impairment and Alzheimer's disease. *J Chem Neuroanat*. 2011;42(2):102-110.
293. Hubert T, Grimal S, Carroll P, Fichard-Carroll A. Collagens in the developing and diseased nervous system. *Cell Mol Life Sci*. 2009;66(7):1223-1238.
294. Chernousov MA, Rothblum K, Stahl RC, Evans A, Prentiss L, Carey DJ. Glypican-1 and alpha4(V) collagen are required for Schwann cell myelination. *J Neurosci*. 2006;26(2):508-517.
295. Heffron DS, Landreth GE, Samuels IS, Mandell JW. Brain-specific deletion of extracellular signal-regulated kinase 2 mitogen-activated protein kinase leads to aberrant cortical collagen deposition. *Am J Pathol*. 2009;175(6):2586-2599.

References

296. Tong Y, Xu Y, Scearce-Levie K, Ptáček LJ, Fu YH. COL25A1 triggers and promotes Alzheimer's disease-like pathology in vivo. *Neurogenetics*. 2010;11(1):41-52.
297. Gold M, Dolga AM, Koepke J, et al. α 1-antitrypsin modulates microglial-mediated neuroinflammation and protects microglial cells from amyloid- β -induced toxicity. *J Neuroinflammation*. 2014;11:165.
298. Jesse S, Lehnert S, Jahn O, et al. Differential sialylation of serpin A1 in the early diagnosis of Parkinson's disease dementia. *PloS one*. 2012;7(11):e48783-e48783.
299. Nielsen HM, Minthon L, Londos E, et al. Plasma and CSF serpins in Alzheimer disease and dementia with Lewy bodies. *Neurology*. 2007;69(16):1569-1579.
300. Saman S, Kim W, Raya M, et al. Exosome-associated tau is secreted in tauopathy models and is selectively phosphorylated in cerebrospinal fluid in early Alzheimer disease. *J Biol Chem*. 2012;287(6):3842-3849.
301. Asai H, Ikezu S, Tsunoda S, et al. Depletion of microglia and inhibition of exosome synthesis halt tau propagation. *Nat Neurosci*. 2015;18(11):1584-1593.
302. Rajendran L, Honsho M, Zahn TR, et al. Alzheimer's disease beta-amyloid peptides are released in association with exosomes. *Proc Natl Acad Sci U S A*. 2006;103(30):11172-11177.
303. Zenaro E, Pietronigro E, Della Bianca V, et al. Neutrophils promote Alzheimer's disease-like pathology and cognitive decline via LFA-1 integrin. *Nat Med*. 2015;21(8):880-886.
304. Xu J, Patassini S, Rustogi N, et al. Regional protein expression in human Alzheimer's brain correlates with disease severity. *Communications Biology*. 2019;2(1):43.
305. Zhang X, Yang W, Wang J, Meng Y, Guan Y, Yang J. FAM3 gene family: A promising therapeutic target for NAFLD and type 2 diabetes. *Metabolism*. 2018;81:71-82.
306. Hasegawa H, Liu L, Tooyama I, Murayama S, Nishimura M. The FAM3 superfamily member ILEI ameliorates Alzheimer's disease-like pathology by destabilizing the penultimate amyloid- β precursor. *Nature Communications*. 2014;5(1):3917.
307. Nakano M, Mitsuishi Y, Liu L, et al. Extracellular Release of ILEI/FAM3C and Amyloid- β Is Associated with the Activation of Distinct Synapse Subpopulations. *J Alzheimers Dis*. 2021;80(1):159-174.
308. Liu L, Watanabe N, Akatsu H, Nishimura M. Neuronal expression of ILEI/FAM3C and its reduction in Alzheimer's disease. *Neuroscience*. 2016;330:236-246.
309. Bonneh-Barkay D, Bissel SJ, Wang G, et al. YKL-40, a marker of simian immunodeficiency virus encephalitis, modulates the biological activity of basic fibroblast growth factor. *Am J Pathol*. 2008;173(1):130-143.
310. Shao R, Hamel K, Petersen L, et al. YKL-40, a secreted glycoprotein, promotes tumor angiogenesis. *Oncogene*. 2009;28(50):4456-4468.
311. Choi J, Lee HW, Suk K. Plasma level of chitinase 3-like 1 protein increases in patients with early Alzheimer's disease. *J Neurol*. 2011;258(12):2181-2185.

References

312. Vinther-Jensen T, Börnsen L, Budtz-Jørgensen E, et al. Selected CSF biomarkers indicate no evidence of early neuroinflammation in Huntington disease. *Neurol Neuroimmunol Neuroinflamm.* 2016;3(6):e287-e287.
313. Baldacci F, Lista S, Cavedo E, Bonuccelli U, Hampel H. Diagnostic function of the neuroinflammatory biomarker YKL-40 in Alzheimer's disease and other neurodegenerative diseases. *Expert Rev Proteomics.* 2017;14(4):285-299.
314. Goodman L. Alzheimer's disease; a clinico-pathologic analysis of twenty-three cases with a theory on pathogenesis. *J Nerv Ment Dis.* 1953;118(2):97-130.
315. Zhang X, Alshakhshir N, Zhao L. Glycolytic Metabolism, Brain Resilience, and Alzheimer's Disease. *Frontiers in Neuroscience.* 2021;15(476).
316. An Y, Varma VR, Varma S, et al. Evidence for brain glucose dysregulation in Alzheimer's disease. *Alzheimers Dement.* 2018;14(3):318-329.
317. Madeira C, Lourenco MV, Vargas-Lopes C, et al. d-serine levels in Alzheimer's disease: implications for novel biomarker development. *Transl Psychiatry.* 2015;5(5):e561.
318. Campbell NL, Unverzagt F, LaMantia MA, Khan BA, Boustani MA. Risk factors for the progression of mild cognitive impairment to dementia. *Clin Geriatr Med.* 2013;29(4):873-893.
319. Matsuyama SS, Jarvik LF. Hypothesis: microtubules, a key to Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America.* 1989;86(20):8152-8156.
320. Ma RH, Zhang Y, Hong XY, Zhang JF, Wang JZ, Liu GP. Role of microtubule-associated protein tau phosphorylation in Alzheimer's disease. *J Huazhong Univ Sci Technolog Med Sci.* 2017;37(3):307-312.
321. Thues C, Valadas JS, Deaulmerie L, et al. MAPRE2 mutations result in altered human cranial neural crest migration, underlying craniofacial malformations in CSC-KT syndrome. *Scientific Reports.* 2021;11(1):4976.
322. Gong C-X, Liu F, Grundke-Iqbal I, Iqbal K. Dysregulation of protein phosphorylation/dephosphorylation in Alzheimer's disease: a therapeutic target. *J Biomed Biotechnol.* 2006;2006(3):31825-31825.
323. Zhang Q, Ma C, Gearing M, Wang PG, Chin L-S, Li L. Integrated proteomics and network analysis identifies protein hubs and network alterations in Alzheimer's disease. *Acta neuropathologica communications.* 2018;6(1):19-19.
324. Sontag J-M, Sontag E. Protein phosphatase 2A dysfunction in Alzheimer's disease. *Frontiers in Molecular Neuroscience.* 2014;7(16).
325. Sontag E, Hladik C, Montgomery L, et al. Downregulation of Protein Phosphatase 2A Carboxyl Methylation and Methyltransferase May Contribute to Alzheimer Disease

References

- Pathogenesis. *Journal of Neuropathology & Experimental Neurology*. 2004;63(10):1080-1091.
326. Choi JE, Lee JJ, Kang W, et al. Proteomic Analysis of Hippocampus in a Mouse Model of Depression Reveals Neuroprotective Function of Ubiquitin C-terminal Hydrolase L1 (UCH-L1) via Stress-induced Cysteine Oxidative Modifications. *Mol Cell Proteomics*. 2018;17(9):1803-1823.
327. Zheng H, Koo EH. Biology and pathophysiology of the amyloid precursor protein. *Mol Neurodegener*. 2011;6(1):27.
328. Kim J, Yoon H, Basak J, Kim J. Apolipoprotein E in synaptic plasticity and Alzheimer's disease: potential cellular and molecular mechanisms. *Mol Cells*. 2014;37(11):767-776.
329. Liao F, Yoon H, Kim J. Apolipoprotein E metabolism and functions in brain and its role in Alzheimer's disease. *Curr Opin Lipidol*. 2017;28(1):60-67.
330. Konijnenberg E, Tijms BM, Gobom J, et al. APOE ϵ 4 genotype-dependent cerebrospinal fluid proteomic signatures in Alzheimer's disease. *Alzheimer's Research & Therapy*. 2020;12(1):65.
331. Berger M, Cooter M, Roesler AS, et al. APOE4 Copy Number-Dependent Proteomic Changes in the Cerebrospinal Fluid. *J Alzheimers Dis*. 2021;79(2):511-530.
332. Villemagne VL, Pike KE, Ch  telat G, et al. Longitudinal assessment of A β and cognition in aging and Alzheimer disease. *Ann Neurol*. 2011;69(1):181-192.
333. Yu G, Wang L-G, Yan G-R, He Q-Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics*. 2015;31(4):608-609.
334. Tsatsanis A, McCorkindale AN, Wong BX, et al. The acute phase protein lactoferrin is a key feature of Alzheimer's disease and predictor of A β burden through induction of APP amyloidogenic processing. *Molecular Psychiatry*. 2021.
335. Goemaere J, Knoop B. Peroxiredoxin distribution in the mouse brain with emphasis on neuronal populations affected in neurodegenerative disorders. *J Comp Neurol*. 2012;520(2):258-280.
336. Clausen A, Xu X, Bi X, Baudry M. Effects of the superoxide dismutase/catalase mimetic EUK-207 in a mouse model of Alzheimer's disease: protection against and interruption of progression of amyloid and tau pathology and cognitive decline. *J Alzheimers Dis*. 2012;30(1):183-208.
337. Zhang Z, Yan J, Chang Y, ShiDu Yan S, Shi H. Hypoxia inducible factor-1 as a target for neurodegenerative diseases. *Curr Med Chem*. 2011;18(28):4335-4343.
338. Correia SC, Moreira PI. Hypoxia-inducible factor 1: a new hope to counteract neurodegeneration? *J Neurochem*. 2010;112(1):1-12.
339. Mahley RW. Central Nervous System Lipoproteins: ApoE and Regulation of Cholesterol Metabolism. *Arterioscler Thromb Vasc Biol*. 2016;36(7):1305-1315.

References

340. Aroner SA, Yang M, Li J, et al. Apolipoprotein C-III and High-Density Lipoprotein Subspecies Defined by Apolipoprotein C-III in Relation to Diabetes Risk. *Am J Epidemiol*. 2017;186(6):736-744.
341. Thambisetty M. Plasma Apolipoprotein E Levels and Risk of Dementia—You Are the Company You Keep. *JAMA Network Open*. 2020;3(7):e209501-e209501.
342. Koch M, DeKosky ST, Goodman M, et al. Association of Apolipoprotein E in Lipoprotein Subspecies With Risk of Dementia. *JAMA Network Open*. 2020;3(7):e209250-e209250.
343. Koch M, DeKosky ST, Goodman M, et al. High density lipoprotein and its apolipoprotein-defined subspecies and risk of dementia. *J Lipid Res*. 2020;61(3):445-454.
344. Feringa FM, van der Kant R. Cholesterol and Alzheimer's Disease; From Risk Genes to Pathological Effects. *Front Aging Neurosci*. 2021;13(333).
345. Israelsen WJ, Vander Heiden MG. Pyruvate kinase: Function, regulation and role in cancer. *Semin Cell Dev Biol*. 2015;43:43-51.
346. Nakamura T, Watanabe A, Fujino T, Hosono T, Michikawa M. Apolipoprotein E4 (1-272) fragment is associated with mitochondrial proteins and affects mitochondrial function in neuronal cells. *Mol Neurodegener*. 2009;4:35.
347. Perkins M, Wolf AB, Chavira B, et al. Altered Energy Metabolism Pathways in the Posterior Cingulate in Young Adult Apolipoprotein E ϵ 4 Carriers. *J Alzheimers Dis*. 2016;53(1):95-106.
348. Liu H, Harrell LE, Shenvi S, Hagen T, Liu RM. Gender differences in glutathione metabolism in Alzheimer's disease. *J Neurosci Res*. 2005;79(6):861-867.
349. Agarwal S, Sohal RS. DNA oxidative damage and life expectancy in houseflies. *Proc Natl Acad Sci U S A*. 1994;91(25):12332-12335.
350. Mazzanti G, Di Giacomo S. Curcumin and Resveratrol in the Management of Cognitive Disorders: What is the Clinical Evidence? *Molecules*. 2016;21(9):1243.
351. Kaminsky YG, Reddy VP, Ashraf GM, et al. Age-related defects in erythrocyte 2,3-diphosphoglycerate metabolism in dementia. *Aging Dis*. 2013;4(5):244-255.
352. Nóbrega-Pereira S, Fernandez-Marcos PJ, Brioché T, et al. G6PD protects from oxidative damage and improves healthspan in mice. *Nature communications*. 2016;7:10894-10894.
353. Evlice A, Ulusu NN. Glucose-6-phosphate dehydrogenase a novel hope on a blood-based diagnosis of Alzheimer's disease. *Acta Neurologica Belgica*. 2017;117(1):229-234.
354. Carlini V, Verduci I, Cianci F, et al. CLIC1 Protein Accumulates in Circulating Monocyte Membrane during Neurodegeneration. *Int J Mol Sci*. 2020;21(4):1484.
355. Orre M, Kamphuis W, Osborn LM, et al. Isolation of glia from Alzheimer's mice reveals inflammation and dysfunction. *Neurobiol Aging*. 2014;35(12):2746-2760.

References

356. Emrani S, Arain HA, DeMarshall C, Nuriel T. APOE4 is associated with cognitive and pathological heterogeneity in patients with Alzheimer's disease: a systematic review. *Alzheimer's Research & Therapy*. 2020;12(1):141.
357. Tao Q, Ang TFA, DeCarli C, et al. Association of Chronic Low-grade Inflammation With Risk of Alzheimer Disease in ApoE4 Carriers. *JAMA Netw Open*. 2018;1(6):e183597.
358. Agosta F, Vessel KA, Miller BL, et al. Apolipoprotein E epsilon4 is associated with disease-specific effects on brain atrophy in Alzheimer's disease and frontotemporal dementia. *Proc Natl Acad Sci U S A*. 2009;106(6):2018-2022.
359. Kim SH, Fountoulakis M, Cairns N, Lubec G. Protein levels of human peroxiredoxin subtypes in brains of patients with Alzheimer's disease and Down syndrome. *J Neural Transm Suppl*. 2001(61):223-235.
360. Szeliga M. Peroxiredoxins in Neurodegenerative Diseases. *Antioxidants (Basel)*. 2020;9(12).
361. Favrin G, Bean DM, Bilsland E, et al. Identification of novel modifiers of A β toxicity by transcriptomic analysis in the fruitfly. *Scientific Reports*. 2013;3(1):3512.
362. Cheng JS, Dubal DB, Kim DH, et al. Collagen VI protects neurons against Abeta toxicity. *Nature neuroscience*. 2009;12(2):119-121.
363. Akiyama H, Barger S, Barnum S, et al. Inflammation and Alzheimer's disease. *Neurobiol Aging*. 2000;21(3):383-421.
364. Veerhuis R, Nielsen HM, Tenner AJ. Complement in the brain. *Mol Immunol*. 2011;48(14):1592-1603.
365. Dejanovic B, Huntley MA, De Mazière A, et al. Changes in the Synaptic Proteome in Tauopathy and Rescue of Tau-Induced Synapse Loss by C1q Antibodies. *Neuron*. 2018;100(6):1322-1336.e1327.
366. Litvinchuk A, Wan Y-W, Swartzlander DB, et al. Complement C3aR Inactivation Attenuates Tau Pathology and Reverses an Immune Network Deregulated in Tauopathy Models and Alzheimer's Disease. *Neuron*. 2018;100(6):1337-1353.e1335.
367. Toledo JB, Korff A, Shaw LM, Trojanowski JQ, Zhang J. Low levels of cerebrospinal fluid complement 3 and factor H predict faster cognitive decline in mild cognitive impairment. *Alzheimers Res Ther*. 2014;6(3):36.
368. Bonham LW, Desikan RS, Yokoyama JS. The relationship between complement factor C3, APOE ϵ 4, amyloid and tau in Alzheimer's disease. *Acta Neuropathol Commun*. 2016;4(1):65.
369. Finehout EJ, Franck Z, Lee KH. Complement protein isoforms in CSF as possible biomarkers for neurodegenerative disease. *Dis Markers*. 2005;21(2):93-101.
370. Reale M, Kamal MA, Velluto L, Gambi D, Di Nicola M, Greig NH. Relationship between inflammatory mediators, A β levels and ApoE genotype in Alzheimer disease. *Curr Alzheimer Res*. 2012;9(4):447-457.

References

371. Mahoney ER, Dumitrescu L, Moore AM, et al. Brain expression of the vascular endothelial growth factor gene family in cognitive aging and alzheimer's disease. *Molecular Psychiatry*. 2021;26(3):888-896.
372. Lin R, Zhang Y, Yan D, et al. Association of common variants in TOMM40/APOE/APOC1 region with human longevity in a Chinese population. *J Hum Genet*. 2016;61(4):323-328.
373. Fuior EV, Gafencu AV. Apolipoprotein C1: Its Pleiotropic Effects in Lipid Metabolism and Beyond. *International journal of molecular sciences*. 2019;20(23):5939.
374. Polanski M, Anderson NL. A list of candidate cancer biomarkers for targeted proteomics. *Biomark Insights*. 2007;1:1-48.
375. Parker CE, Borchers CH. Mass spectrometry based biomarker discovery, verification, and validation--quality assurance and control of protein biomarker assays. *Mol Oncol*. 2014;8(4):840-858.
376. Younes L, Albert M, Moghekar A, Soldan A, Pettigrew C, Miller MI. Identifying Changepoints in Biomarkers During the Preclinical Phase of Alzheimer's Disease. *Frontiers in Aging Neuroscience*. 2019;11(74).

Fluid Biomarkers and *APOE* Status of Early Onset Alzheimer's Disease Variants: A Systematic Review and Meta-Analysis

Gurjeet Kaur^a, Anne Poljak^{a,b,c}, Nady Braidy^a, John D. Crawford^a,
Jessica Lo^a and Perminder S. Sachdev^{a,d,*}

^aCentre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, NSW, Australia

^bMark Wainwright Analytical Centre, Bioanalytical Mass Spectrometry Facility, University of New South Wales, Sydney, NSW, Australia

^cSchool of Medical Sciences, University of New South Wales, Sydney, NSW, Australia

^dNeuropsychiatric Institute, Euroa Centre, Prince of Wales Hospital, Sydney, NSW, Australia

Accepted 13 March 2020

Abstract.

Background: Numerous studies have reported on cerebrospinal fluid (CSF) and blood biomarkers of Alzheimer's disease (AD); however, to date, none has compared biomarker patterns across the early-onset subtypes, i.e., early onset sporadic AD (EOsAD) and autosomal dominant AD (ADAD), qualitatively and quantitatively.

Objective: To compare the fluid biomarker patterns in early-onset subtypes of AD; EOsAD and ADAD.

Methods: Six scientific databases were searched for peer-reviewed research publications. The total number of individuals used in all the meta-analysis were 2,427, comprised of 1,337 patients and 1,090 controls.

Results: In the subset of EOsAD cases without *APP*, *PSEN1/PSEN2* mutations, CSF A β ₄₂ and tau levels were higher when compared to the EOsAD group as a whole. Prevalence of the *APOE* ϵ 4 allele was more elevated in EOsAD relative to controls, and not significantly elevated in ADAD cases.

Conclusion: Established CSF biomarkers confirmed quantitative differences between variants of EOAD. EOsAD is enriched with *APOE* ϵ 4, but the level is not higher than generally reported in late-onset AD. The results prompt further exploration of the etiopathogenesis of EOsAD, which accounts for ~4–10% of all AD cases, but the reasons for the early onset remain poorly understood.

Keywords: Amyloid- β ₄₂, *APOE* ϵ 4, *APP/PSEN*, early onset Alzheimer's disease, neurodegeneration biomarkers, tau

INTRODUCTION

Despite intensive research over the last two decades, no reliable treatment exists to reverse or impede the onset and progression of Alzheimer's disease (AD), which is usually sub-categorized into two

subtypes based on the age of onset: early onset AD (EOAD) and late-onset AD (LOAD), with age 65 years as the generally accepted cut-off [1]. EOAD accounts for only 5%–10% of all AD cases [2], with <1% being attributed to autosomal dominant AD (ADAD), caused by mutations in one of three genes: amyloid precursor protein (*APP*), presenilin 1 (*PSEN1*), and presenilin 2 (*PSEN2*). ADAD cases have 100% penetrance of a known genetic mutation, with predictable age of symptom onset, and therefore presymptomatic individuals can be studied

*Correspondence to: Professor Perminder Sachdev, MD, PhD, Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, UNSW Medicine, UNSW, Sydney, NSW, 2052, Australia. Tel.: +61(2) 9385 7663; Fax: +61(2) 9385 3645; E-mail: p.sachdev@unsw.edu.au.

several years before symptom onset. ADAD cases generally represent a “purer” version of AD with fewer concomitant pathologies and hence confounding variables [3]. Whereas, early onset sporadic AD (EOsAD) is rarely studied in its own right and is, therefore, a particularly important subset as it presents an opportunity to either discover new lifestyle risk factors or explore novel mutations responsible for the earlier age of onset. Genetic variations, in particular that of the Apolipoprotein E (*APOE*) gene, play a significant role in LOAD but are poorly understood in EOsAD. ADAD and EOsAD can be similar pathologically and phenotypically, with the primary distinction being the presence of autosomal dominant mutation in the former and possibly risk genes and environmental/lifestyle factors in the latter. However, there are several reasons why biomarkers may differ for the two disorders: 1) That they are pathologically similar does not imply that the pathology is exactly the same. Considering the difference in etiology, it is possible that there are subtle differences in the pathology, e.g., the relative amyloid and tau burden, the degree of inflammation, oxidative stress, and synaptic loss, which may be reflected in quantitative differences in biomarkers; 2) There are good reasons to believe that over-production of amyloid- β_{42} is an earlier and defining feature of ADAD pathophysiology, but one could argue that this may not be the central disturbance in EOsAD, which would be reflected in the biomarkers, in particular $A\beta$ levels; and 3) The rate of development and progression of pathology is different in the two disorders such that at a particular stage of the disease (as identified by cognitive deficits), the biomarkers may differ quantitatively. We therefore started with the hypothesis that the biomarkers would be different. However, had we set up a null hypothesis, i.e., there would be no difference in the biomarkers for the two disorders at the similar levels of cognitive deficit, our methodology or conclusions would not have been different.

The failure of numerous clinical trials suggests that a focus on presymptomatic biomarkers with a view to prevention may be a better option than treatment once pathology has significantly progressed.

Absence of comparative studies reporting pathological and phenotypical difference/s in ADAD and EOsAD makes it difficult to know whether there are qualitatively or quantitatively similar or different biomarker patterns in these genetically different AD subtypes. We therefore performed a systematic review and meta-analysis on variants of EOAD to

examine the shared as well as unique factors that influence subtypes of EOAD.

METHODS

Search strategy

This systematic review and meta-analysis followed Preferred Reporting Items in PRISMA guidelines [4] and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [5]. Literature searches were performed for papers published between January 1, 1988, and June 1, 2019, using six major scientific databases: Medline Ovid, PubMed, EMBASE, PsycINFO, Web of Science, and Scopus (Fig. 1, Supplementary Tables 2 and 3).

Data extraction

Inclusion/Exclusion criteria

Included in the review were papers reporting on quantified biomarkers (proteins/peptides) in cerebrospinal fluid (CSF) and/or blood (serum or plasma) in patients with EOsAD or ADAD and corresponding controls. These searches included combinations of the following Boolean terms: [blood OR serum OR plasma OR cerebrospinal fluid OR CSF] AND [autosomal dominant OR dominant inheritance OR early onset OR young-onset Alzheimer’s disease] and were run in all six databases. Exclusion criteria were: 1) studies purely on LOAD or on any AD subtype with biomarker data on tissues other than blood, serum, plasma, or CSF, studies reporting purely on non-protein biomarkers such as mRNA (Q-PCR) or studies which use non-quantitative observations (e.g., SDS-PAGE without identifying or quantifying specific protein changes); 2) studies purely on animal models, cell culture techniques, immunohistochemistry, genotyping, imaging (PET, FDG), laboratory tests such as blood cell counts, and identification of biomarkers from cellular blood fractions such as peripheral blood cells; 3) articles reporting data following treatment, therapy, intervention, toxicology or human *in vivo* challenge studies, with either endogenous or synthetic/pharmaceutical compounds and without baseline data, and 4) studies on diseases with early onset dementia as a component of the phenotype but of predominantly different etiology (e.g., vascular). Given the limited volume of literature on this topic, no studies were excluded based on cohort size and case studies were included in the systematic review, though not in the meta-analysis. However,

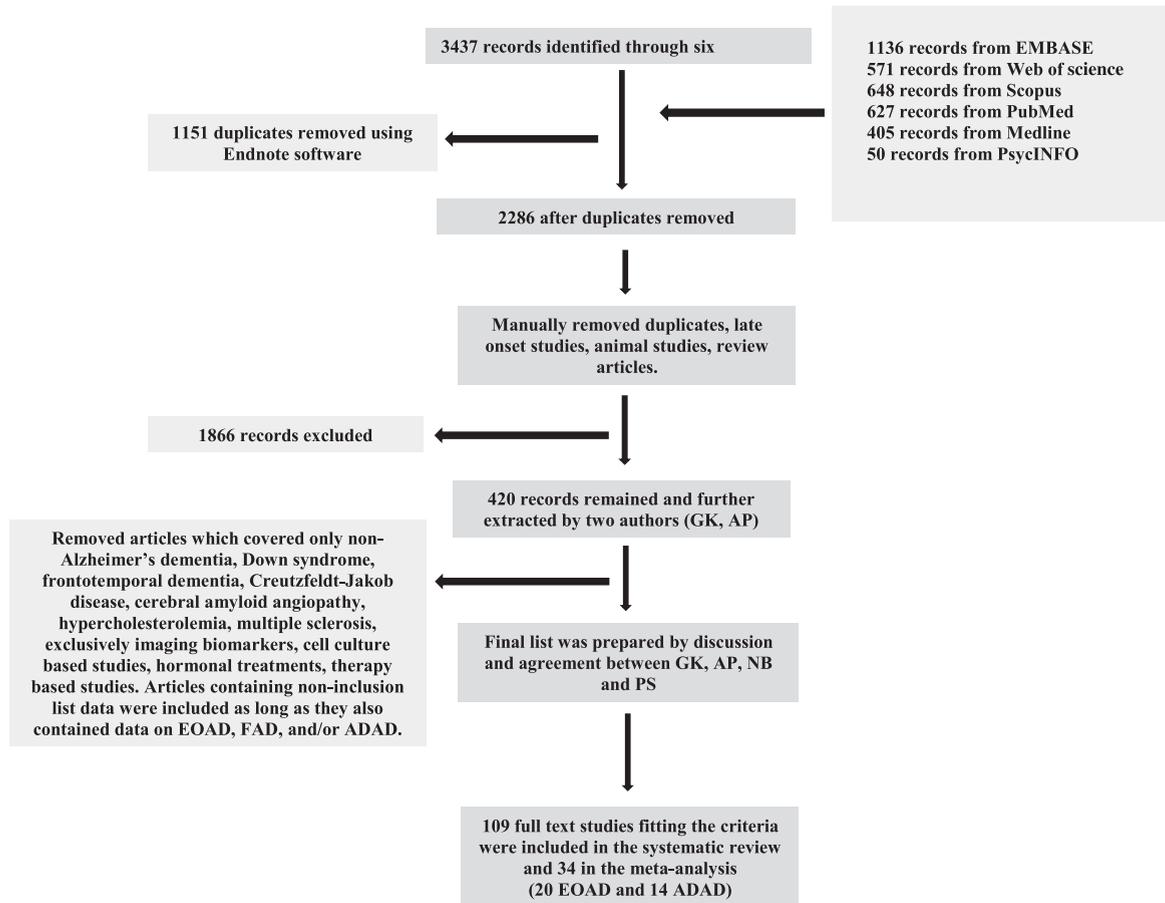


Fig. 1. Flow chart depicting the selection process of records found using six scientific databases and elimination of non-relevant articles.

only cross-sectional studies were available in sufficient numbers to allow meta-analysis.

The meta-analysis included 34 studies, 20 for EOsAD and 14 for ADAD (which included both symptomatic and asymptomatic mutation carriers, sMC and aMC, respectively). By definition, the EOsAD individuals received a clinical diagnosis of AD with onset before 65 years, and the majority of them had no family history of EOAD. Genetic testing was generally not performed to eliminate *APP*, *PSEN1*, and *PSEN2* mutations; however, in 4 of the 20 EOsAD studies, genetic testing was indeed carried out, and meta-analyses were also performed on these separately. The ADAD individuals had a similar age of onset, a positive family history of ADAD, and carried pathogenic mutation/s to one of the following genes: *APP*, *PSEN1*, and *PSEN2*. Across all studies, biomarker level changes were expressed as standardized mean differences (SMD), calculated relative to levels in cognitively healthy age-matched controls.

Five articles identified as meeting inclusion criteria were published on participants from the Dominantly Inherited Alzheimer's Network (DIAN) cohort. After discussion with DIAN authors, and in order to avoid duplication of data on the same participants, we included only the most recently published DIAN article [6], excluding the previous four from the meta-analysis where the biomarker data was duplicated [7–10]. For articles with insufficient data, or data presented in an inaccessible manner, we emailed authors requesting additional information. The majority either did not respond or no longer had access to data. One author provided data which is included in our analyses [11].

While we did not use assay type as a selection criterion (either inclusion or exclusion) for the studies reported here. However, the majority (>95%) of studies used commercial kits which have been manufacturer validated. In particular, of the 27 studies which reported assaying A β and/or

tau variants, 26 used the Innogenetics (Fujirebio-Europe) xMAP technology-based ELISA kits (Supplementary Tables 2 and 3). This fortuitous “standardization” is likely to minimize inter-laboratory variation for the meta-analysis results of A β and tau variants reported here.

Meta-analyses

Meta-analyses were conducted using the procedure, *metan*, within the statistical program Stata, version 15.0, using the random-effects model and the inverse-variance weighting of studies. The type-one error rate was set at 0.05. Meta-analyses were performed if there were at least two independent published studies on a biomarker, but in most cases included 3–13 studies.

For studies with continuous biomarker level as the outcome variable, data input into the statistical program were the number of cases, means and standard deviations (SDs) of the biomarker for each of the AD groups and their corresponding control groups. The output of this program expresses the levels of biomarkers as the SMDs, equal to the difference in concentrations between the particular AD group and its control group, divided by the pooled SDs.

For studies with a binary outcome measure (such as $\epsilon 4$ carrier status), data input were the natural logarithms of the odds ratios (log ORs) and their standard errors. These were calculated as follows. If a and b are the number of $\epsilon 4$ carriers in the AD and control groups respectively, and c and d are the corresponding numbers of non-carriers in those groups, then the odds ratio is given by $OR = a*d/b*c$. The standard errors of each log OR are calculated as the square root of $(1/a + 1/b + 1/c + 1/d)$.

The random-effects model was used for all of the meta-analyses based on the assumption that each study produces estimates of the true effects for that particular study population, but that the study populations vary across individual studies due to differences in populations, study design, and assay platforms [12]. The pooled effect sizes produced by meta-analyses using the random-effects model are the estimates of the mean effect sizes for all potential studies satisfying the selection criteria of the review. Publication bias was assessed using the Stata command *metabias*, which was employed to produce funnel plots and to perform the Egger’s test to check for bias, for meta-analyses comprising ≥ 8 studies. For meta-analyses comprising fewer individual studies, it was not possible to draw reliable conclusions

either from the visual inspection of the funnel plots or from the Egger’s test [13]. Comparison of pooled effects from pairs of meta-analyses was achieved using the Stata procedure, *metareg*.

Two articles [14, 15] presented data as median and range, and in these cases, we substituted median for mean and derived SD from range using the following approach: Range/4 (if $n < 70$) or Range/6 (if $n > 70$) [16]. For two articles, mean and SD values were estimated from bar graphs [9, 17]. Two articles reported serum neurofilament light chain (NfL) in ADAD individuals [18, 19], although, for one of them [18], we could not get access to the data and were therefore unable to perform a meta-analysis on serum NfL. A significant share of the ADAD data came from two large cohorts: the DIAN cohort and the Alzheimer’s Prevention Initiative (API) cohort representing a large pedigree living in the state of Antioquia in Colombia, South America. The DIAN cohort includes carrier and non-carrier (NC) family members with various ADAD mutations, while the Colombian kindred likely descended from a single individual and carried the E280A mutation in the *PSEN1* gene.

Data availability statement

Data are available to qualified investigators on request to the corresponding author. Data will be shared at the request of other investigators for purposes of replicating procedures and results.

RESULTS

Protein/peptide biomarkers

Early onset Alzheimer’s disease (EOAD)

We identified 34 eligible published studies on EOAD, comprising 20 studies on EOsAD and 14 on ADAD. The results for established biomarkers (A β_{42} , T-tau, and P-tau) were as follows: All variants of EOAD had lower CSF A β_{42} and higher CSF T-tau and P-tau (Table 1, Figs. 2–4), relative to their respective comparison groups. The majority of EOAD biomarkers were identified in CSF, and few studies of blood biomarkers were replicated frequently enough to allow meta-analysis.

In addition to established biomarkers, quantitative data on several biomarkers of neurodegeneration and inflammation were reported in 2 to 4 studies

Table 1

A) Pooled standardized mean difference (pooled SMD), I-squared (I^2), and significance (p) are shown with $p \leq 0.05$ accepted as statistically significant. B) Meta-analysis of *APOE* $\epsilon 4$ carriers in early onset AD pooled odds ratio (pooled OR), I-squared (I^2), and significance (p) are shown with $p \leq 0.05$ accepted as statistically significant (random effects model is represented in the tables). Meta-analysis of CSF and serum/plasma biomarkers of EOAD and ADAD

A	Biomarker	No. of independent studies	EOAD (total numbers)	Controls (total numbers)	Meta-Analysis Random effects model
(i): Early onset sporadic AD (EOsAD)	CSF A β_{42}	8	266	211	Pooled SMD = -1.83, p < 0.001 , $I^2 = 80.9\%$
	CSF T-tau	10	362	247	Pooled SMD = 1.24, p < 0.001 , $I^2 = 80.6\%$
	CSF P-tau	13	375	321	Pooled SMD = 1.39, p < 0.001 , $I^2 = 90\%$
	CSF P-tau ₁₈₁ & ₁₉₉	4	128	142	Pooled SMD = 1.17, p = 0.018 , $I^2 = 92.2\%$
	CSF P-tau ₁₈₁	3	95	102	Pooled SMD = 0.80, $p = 0.11$, $I^2 = 90.5\%$
	CSF NfL	2	72	59	Pooled SMD = 1.42, p < 0.001 , $I^2 = 0.0\%$
	CSF IL6	2	17	39	Pooled SMD = -3.56, $p < 0.385$, $I^2 = 95.2\%$
	CSF IgG	4	63	62	Pooled SMD = -0.33, $p = 0.075$, $I^2 = 0.0\%$
	CSF albumin	3	52	39	Pooled SMD = -0.94, $p = 0.241$, $I^2 = 90.9\%$
	Serum/Plasma albumin	3	52	39	Pooled SMD = 1.66, $p = 0.351$, $I^2 = 96.8\%$
	Serum/Plasma IgG	4	63	62	Pooled SMD = 1.8, p = 0.037 , $I^2 = 93.1\%$
(ii): EOAD with absence of AD mutations confirmed	CSF A β_{42}	4	136	112	Pooled SMD = -2.18, p < 0.001 , $I^2 = 0.0\%$
	CSF T-tau	3	103	91	Pooled SMD = 1.66, p < 0.001 , $I^2 = 85.5\%$
	CSF P-tau	4	136	117	Pooled SMD = 1.45, p < 0.001 , $I^2 = 77.7\%$
(iii): EOAD with no genetic testing reported	CSF A β_{42}	4	130	99	Pooled SMD = -1.45, p < 0.001 , $I^2 = 87.5\%$
	CSF T-tau	7	259	156	Pooled SMD = 1.05, p < 0.001 , $I^2 = 78.1\%$
	CSF P-tau	8	206	164	Pooled SMD = 1.24, p < 0.001 , $I^2 = 91.7\%$
(iv): ADAD All MCs	CSF A β_{42}	12	480	441	Pooled SMD = -2.11, p < 0.001 , $I^2 = 94.0\%$

(Continued)

Table 1
(Continued)

A	Biomarker	No. of independent studies	EOAD (total numbers)	Controls (total numbers)	Meta-Analysis Random effects model
sMCs	CSF A β ₄₂	10	303	344	Pooled SMD = -2.36, p < 0.001 , I ² = 95.3%
aMCs	CSF A β ₄₂	3	81	90	Pooled SMD = -1.64, <i>p</i> = 0.055, I ² = 91.8%
All MCs	CSF A β ₄₀	3	81	120	Pooled SMD = -0.81, p = 0.044 , I ² = 79.4%
All MCs	CSF T-tau	10	318	288	Pooled SMD = 1.53, p < 0.001 , I ² = 83.5%
sMCs	CSF T-tau	3	16	44	Pooled SMD = 2.97, p < 0.001 , I ² = 23.5%
aMCs	CSF T-tau	5	41	55	Pooled SMD = 1.67, p = 0.003 , I ² = 75.9%
All MCs	CSF P-tau	5	283	267	Pooled SMD = 1.63, p < 0.001 , I ² = 75.6%
All MCs	Plasma A β ₄₂	2	164	115	Pooled SMD = 0.69, p < 0.001 , I ² = 52.5%
B	Biomarker	No. of independent studies	Patients APOE ϵ 4+ Patients APOE ϵ 4- (%enrichment)*	Controls APOE ϵ 4+ Controls APOE ϵ 4- (%enrichment)*	Odds Ratio
EOsAD	APOE ϵ 4	4	60 81 (43%)	35 83 (30%)	Pooled OR = 1.72 p = 0.044, I ² = 0.0%
ADAD	APOE ϵ 4	3	17 43 (28%)	25 62 (29%)	Pooled OR = 1.08, <i>p</i> = 0.838, I ² = 0.0%

*LOAD meta-analyses report APOE ϵ 4 allele enrichment in the 13.7% – 40% range. (i) Early onset sporadic AD (EOsAD); defined as the age of onset <65 years and no family history of AD. (ii) EOsAD with the absence of APP, pSEN1, or PSEN2 mutations confirmed by genetic testing. (iii) EOsAD with no genetic testing reported (diagnosis based only on the absence of family history). (iv) Autosomal dominant AD (ADAD); EOAD with the carrier of at least one of APP, PSEN1, or PSEN2 mutations confirmed by genetic testing.

each (Table 1). These were CSF NfL, IL-6, IgG, albumin, serum albumin, and IgG (Table 1, Fig. 2). Of these non-classical biomarkers (i.e., protein/peptide biomarkers other than A β ₄₂, T-tau, and P-tau), the only statistically significant results were for CSF NfL (pooled SMD = 1.42) and serum/plasma IgG (pooled SMD = 1.80) in the subset of all those with EOsAD.

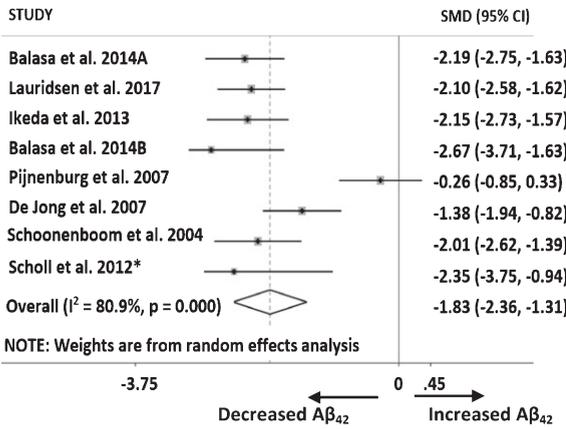
Most of the meta-analyses had high levels of heterogeneity (I² > 75%), reflecting a large amount of scatter around the pooled SMD axis. However, the direction of change, if not its degree, was consistent in >80% of established biomarker studies. Methodological differences across studies likely accounted for this heterogeneity, and therefore the

random-effects model assumptions were the most appropriate [12].

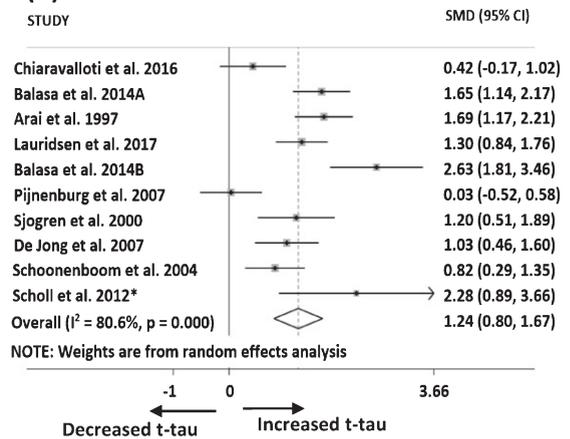
Since this model permits small studies to contribute to outcomes, we explored such publication bias using funnel plots and Egger's test. Funnel plots of 5 biomarkers were performed as they had 8 or more individual studies each (Fig. 5). Although visual inspection of the funnel plots did not give the impression of perfect symmetry around the vertical axis, the results from Egger's test implied that there was no statistically significant systematic relationship between the results of each study and its size.

Comparisons of meta-analysis results of established CSF biomarkers were performed between pairs of patient groups versus controls to examine whether

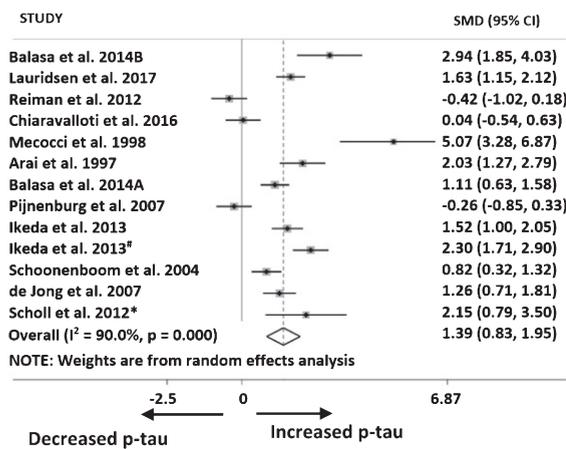
(A)



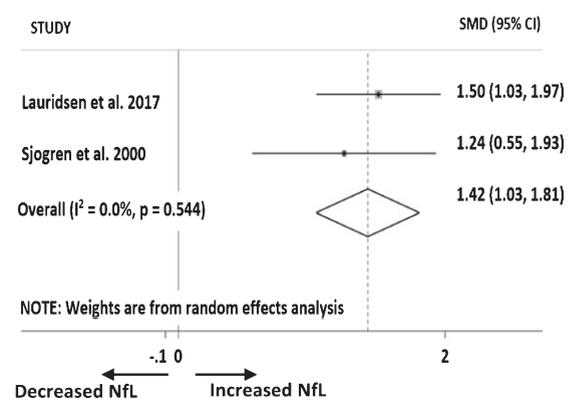
(B)



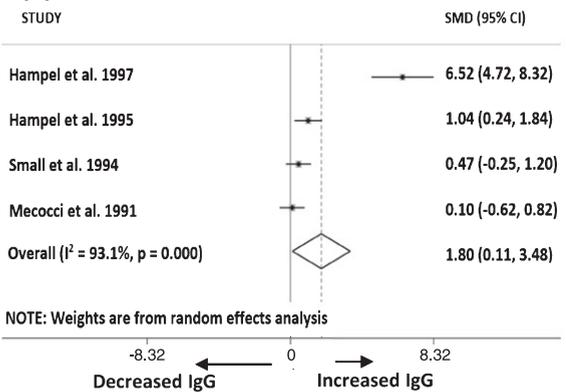
(C)



(D)



(E)



(F)

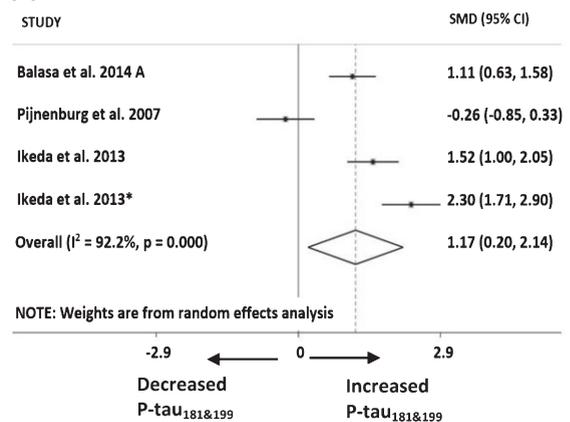


Fig. 2. (Continued)

classical biomarker profiles of EOsAD and subsets of EOsAD are significantly different to those of ADAD (Supplementary Table 1). All the results of the meta-

regression analyses were not significant, except for T-tau in EOsAD versus sMCs of ADAD, suggesting that the classical biomarker profile of EOsAD is

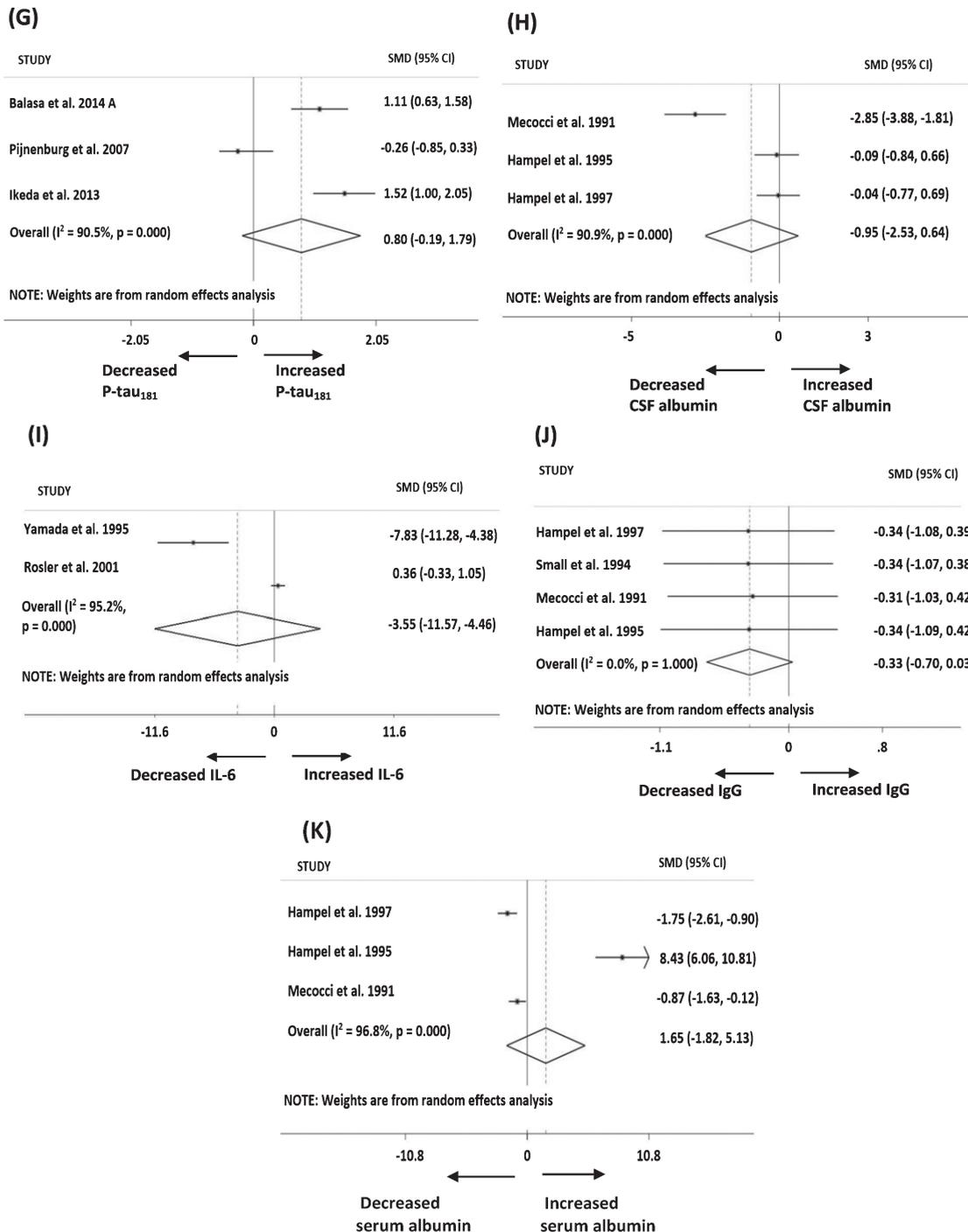


Fig. 2. Forest plots of early onset sporadic Alzheimer’s disease patients versus controls. A) CSF A β_{42} , SMD $p < 0.001$; B) CSF t-tau, SMD $p < 0.001$; C) CSF p-tau, SMD $p < 0.001$; D) CSF NfL, SMD $p < 0.001$; E) serum/plasma IgG, SMD $p = 0.037$; F) CSF P-tau $_{181}$ & $_{199}$, SMD $p = 0.018$; G) CSF P-tau $_{181}$ SMD $p = 0.011$; H) CSF albumin, SMD $p = 0.241$; I) CSF IL-6, SMD $p = 0.385$; J) CSF IgG, SMD $p = 0.075$; K) serum albumin, SMD $p = 0.351$. *values estimated from bar graph. p value represents the statistical significance of SMD test. References for the manuscripts represented in these forest plots are shown in the supplementary section where the full list of manuscripts used for meta-analysis appear in Supplementary Tables 2 and 3. Ikeda et al, 2013 [50] appears twice in panel C and F (Ikeda et al 2013 and Ikeda et al 2013*) as this study analyzed two different isoforms of P-tau, i.e., P-tau $_{181}$ and P-tau $_{199}$, respectively. #Ikeda et al 2013- detected level of tau199 biomarker in panel C and F.

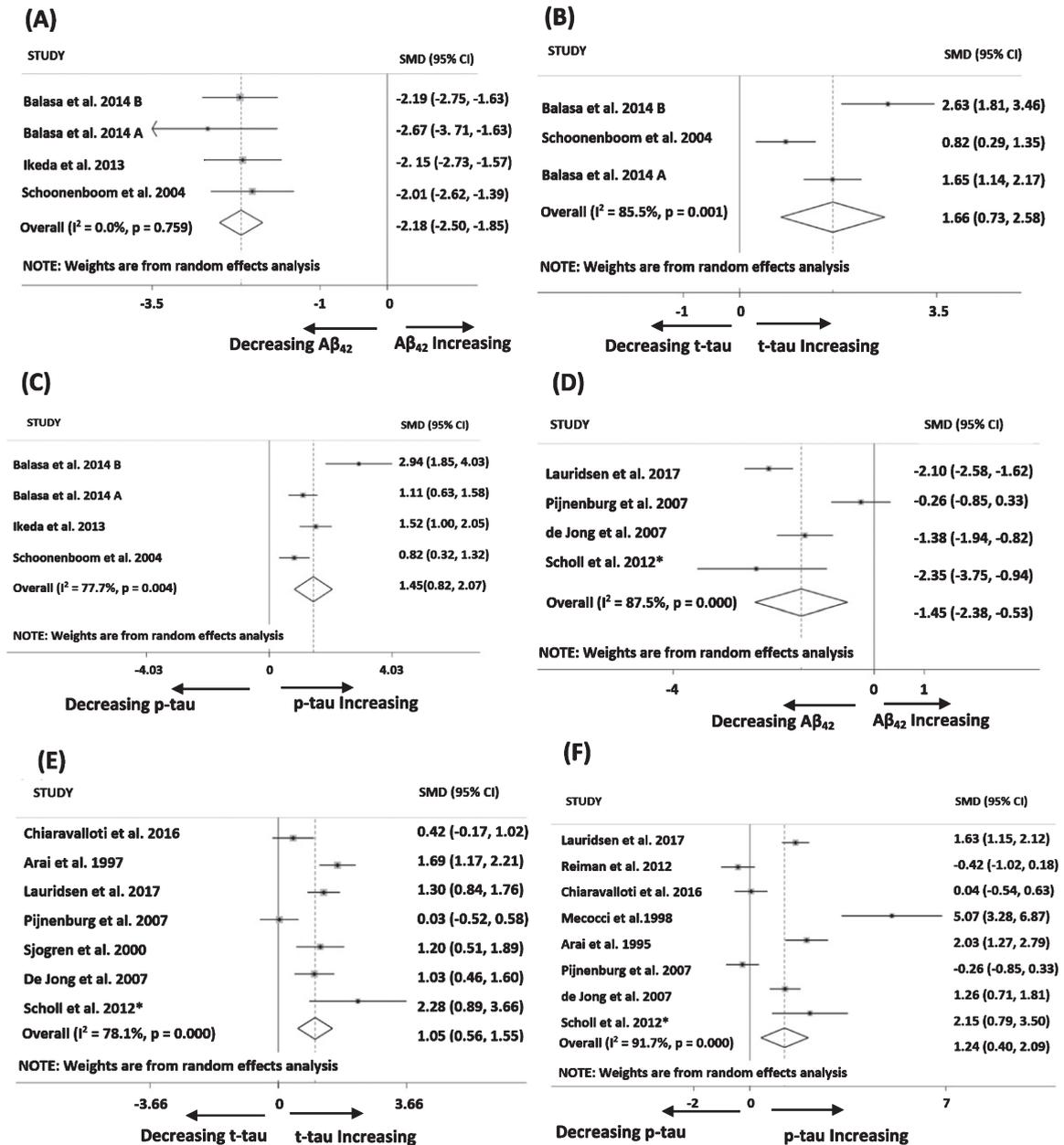


Fig. 3. Forest plots of early onset sporadic Alzheimer’s disease patients (tested negative for *APP/PSEN* mutation carriers) versus normal control individuals. A) CSF A β_{42} (pg/ml); B) CSF t-tau (pg/ml); C) CSF p-tau (pg/ml) and early onset sporadic Alzheimer’s disease patients (no *APP/PSEN* mutation test has been performed) versus normal control individuals; D) CSF A β_{42} (pg/ml); E) CSF t-tau (pg/ml); F) CSF p-tau (pg/ml). *values estimated from bar graph.

not significantly different to the ADAD even though the majority of EOAD individuals have no family history of early onset AD, and some articles on EOAD have even eliminated the possibility of *APP*, *PSEN1*, or *PSEN2* spontaneous mutation/s by genetic testing.

Early onset sporadic Alzheimer’s disease (EOAD)

The 20 EOAD studies collectively included 11 biomarkers (Table 1(i), and Fig. 2) which had been assayed in at least two independent studies (majority >3) and on which meta-analyses could be performed.

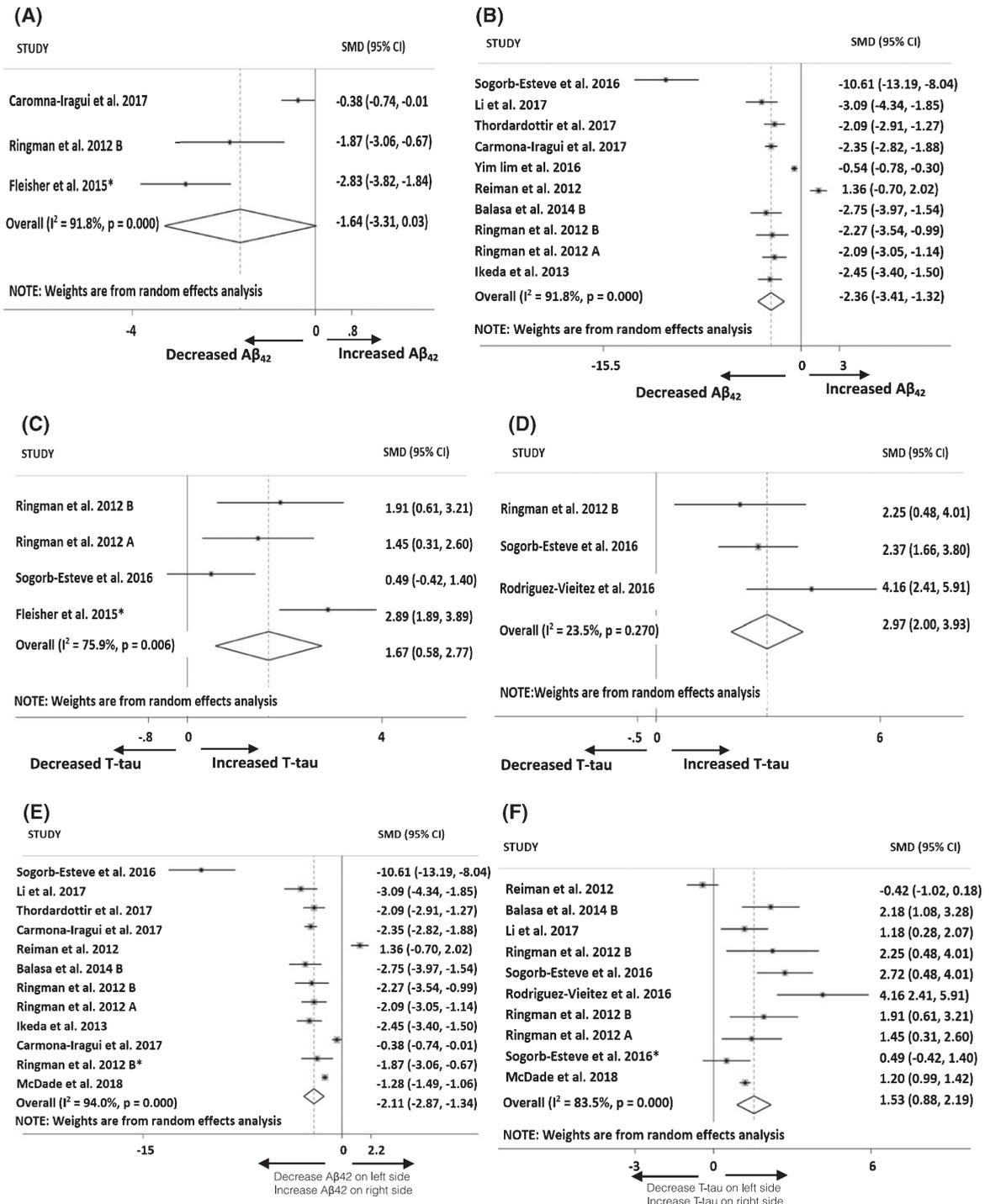


Fig. 4. (Continued)

Significantly lower CSF levels of $A\beta_{42}$ and higher levels of T-tau, P-tau and specific variants of P-tau: P-tau₁₈₁ and ₁₉₉ were observed in EOAD versus control (Table 1(i), Fig. 2). Additional CSF biomarkers, with sufficient numbers of articles (≥ 2 articles)

to allow meta-analysis, included NfL, IgG, IL6, and albumin (Table 1(i), Fig. 2), of which only NfL had a significantly higher level, while the others had no significant difference relative to controls. In serum, only albumin and IgG were assayed in sufficient numbers

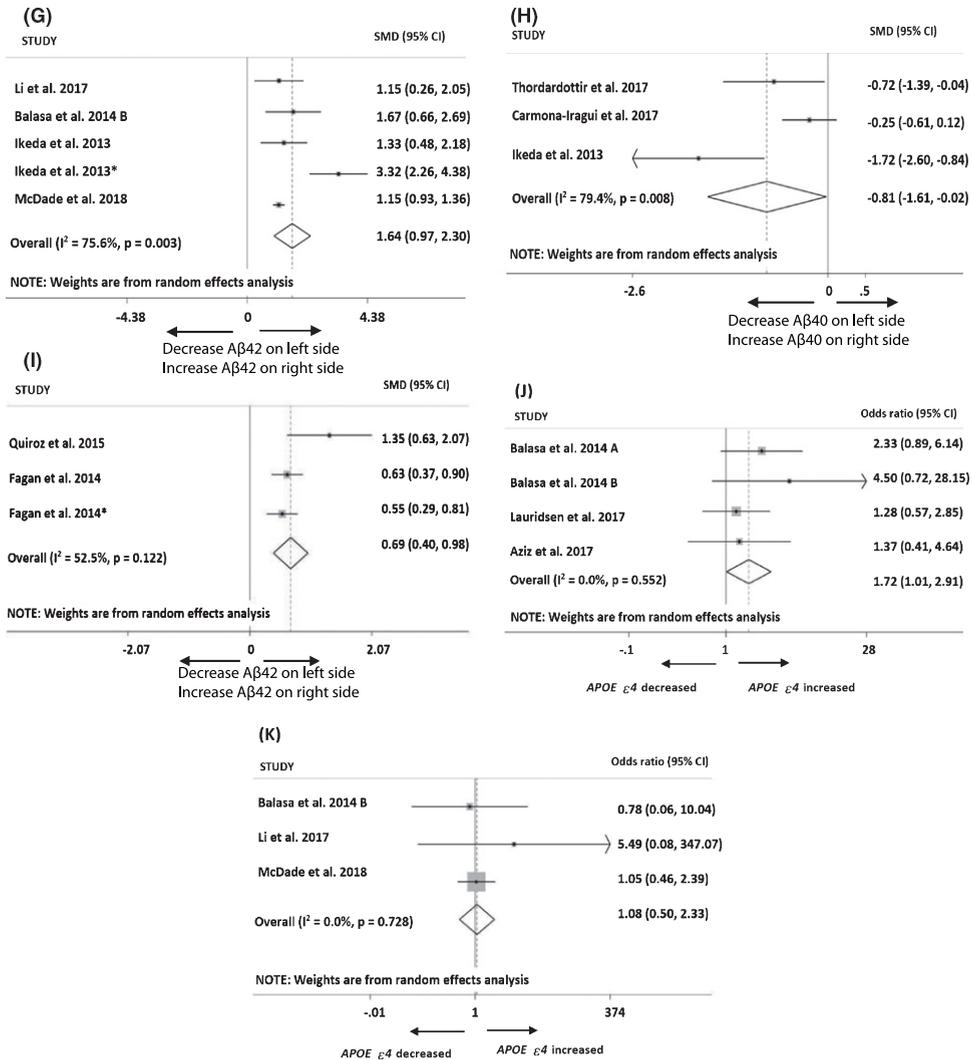


Fig. 4. Forest plots of autosomal dominant Alzheimer’s disease patients versus normal control individuals. A) CSF Aβ₄₂ asymptomatic mutation carriers, SMD $p < 0.001$; B) CSF Aβ₄₂ symptomatic mutation carriers, SMD $p < 0.001$; C) CSF t-tau asymptomatic mutation carriers, SMD $p = 0.003$; D) CSF t-tau symptomatic mutation carriers, SMD $p < 0.001$; E) CSF Aβ₄₂ all mutation carriers, SMD $p < 0.001$; F) CSF T-tau all mutation carriers, SMD $p < 0.001$; G) CSF P-tau all mutation carriers, SMD $p < 0.001$; H) CSF Aβ₄₀ all mutation carriers, SMD $p < 0.044$; I) Plasma Aβ₄₂ all mutation carriers, SMD $p < 0.001$; APOE ε4 allele enrichment in (J) early onset sporadic Alzheimer’s disease, SMD $p = 0.044$; K) autosomal dominant Alzheimer’s disease patients relative to their normal controls SMD $p = 0.838$. *values estimated from bar graph. Fagan et al., 2014 [8] appears twice in panel I (Fagan et al 2014 and Fagan et al 2014*) as this study analyzed Aβ₄₂ in symptomatic mutation carriers and asymptomatic mutation carriers respectively. Ikeda et al., 2013 [50] appears twice in panel G (Ikeda et al 2013 and Ikeda et al 2013*) as this study analyzed two different isoforms of P-tau, i.e., P-tau₁₈₁ and P-tau₁₉₉, respectively.

of studies to allow meta-analysis and only IgG had a statistically significant result (Table 1(i), Fig. 2E), being increased in AD relative to controls.

Of the 20 studies on EOAD, a subset of 4 was identified in which *APP*, *PSEN1*, or *PSEN2* gene mutations had definitively been ruled out by genetic testing (Table 1(ii), Fig. 3A–C). A separate subset of 16 studies did not report confirmation of the absence of these mutations by genetic testing (Table 1(iii),

Fig. 3D–F). Meta-analyses were performed separately on these two sets of studies. As shown in Table 1, pooled SMD values for Aβ₄₂ were lower in the group in which the absence of mutations had been confirmed by genetic testing, while those for both T-tau and P-tau were higher in this group. However, none of these differences were statistically significant when examined using the Stata procedure, *metareg*.

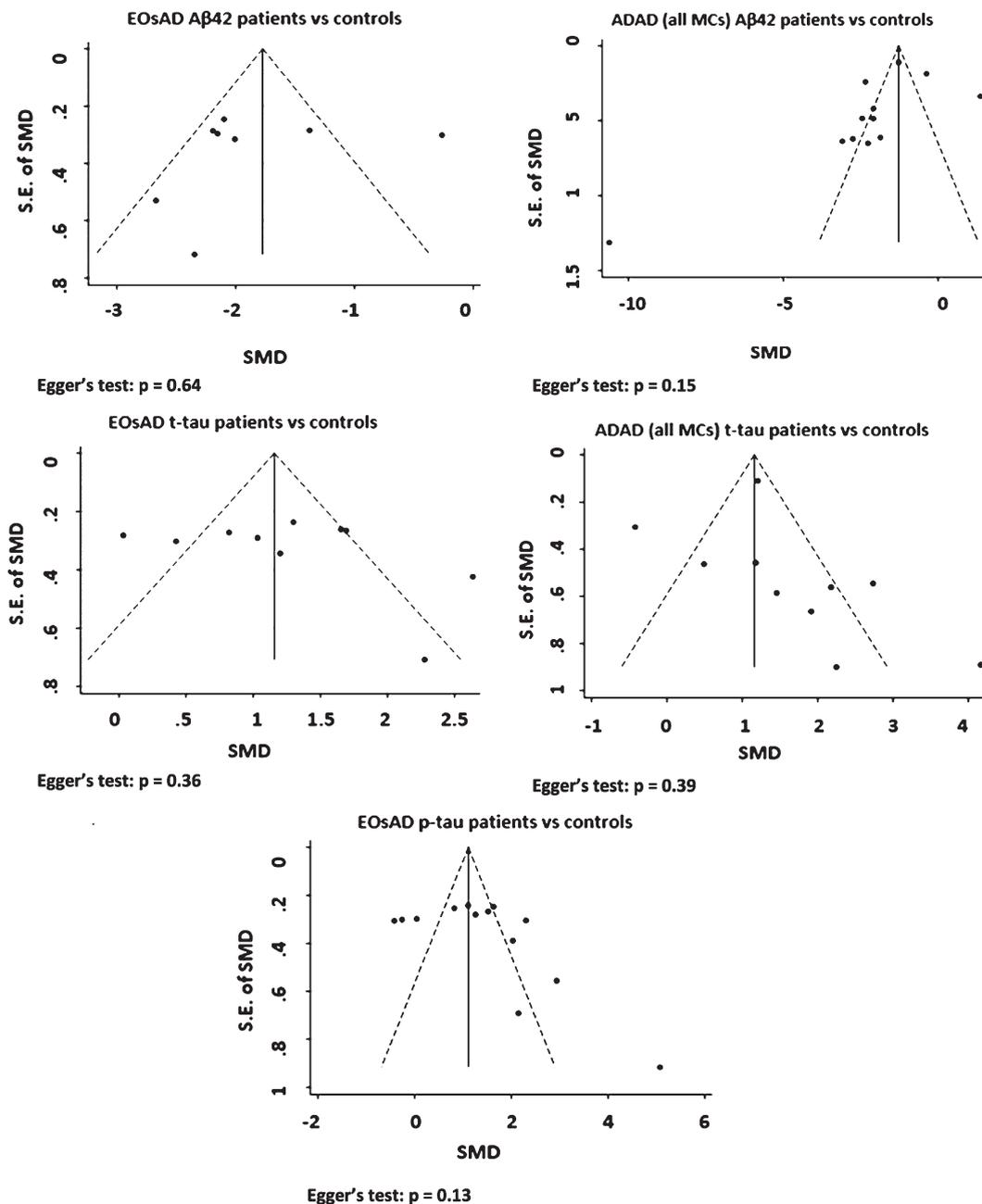


Fig. 5. Funnel plots and Egger's test for publication bias were performed on all meta-analyses which included 8–13 individual studies. For meta-analyses comprising fewer individual studies it is harder to make conclusions about distribution around the SMD axis, and the p -value from the Egger's test is unreliable, so funnel plots are not included for meta-analyses with lower study numbers.

Autosomal dominant Alzheimer's disease (ADAD)

Individuals with a family history of AD and who also carry an AD-related mutation (*APP/PSEN*) are by definition autosomal dominant and may be identified at the aMC stage or at the sMC stage. To

maximize statistical power, we initially performed a meta-analysis on all mutation carriers (all MCs), regardless of stage and subsequently stratified by symptom status (aMCs or sMCs) (Table 1(iv), Fig. 4). AD versus control comparisons for the established biomarkers were all statistically significant in all

MCs, with lower CSF A β_{42} and A β_{40} , higher CSF T-tau, and P-tau and higher plasma A β_{42} for AD (Table 1, Fig. 4). Following stratification by symptom status, CSF A β_{42} and T-tau were significantly lower and higher respectively, in sMCs versus controls (Table 1(iv), Fig. 4B, D). In aMCs, CSF A β_{42} also trended downwards, but pooled SMD was not as low as in symptomatic carriers and did not reach statistical significance (Table 1(iv), Fig. 4A). Level of T-tau significantly increased in aMCs; however, pooled SMD was not as high as in sMCs and results were not statistically significant (Table 1(iv), Fig. 4C). No replicate studies of other biomarkers of AD pathology were reported in ADAD cohorts.

APOE genotype status

APOE $\epsilon 4$ allele enrichment in EOAD groups

A significantly higher proportion of EOAD participants were carriers of the *APOE $\epsilon 4$* allele relative to their controls (Table 1B, Fig. 4J). There was no significant difference in *APOE $\epsilon 4$* prevalence between ADAD and controls (Table 1B, Fig. 4K).

DISCUSSION

We performed a comprehensive meta-analytic review of all currently available protein/peptide CSF and blood biomarkers of early onset AD subtypes and explored these early onset AD phenotypes in the context of quantified biomarkers. We found that the hallmark proteins (A β and tau) showed similar patterns in the CSF, both qualitatively and quantitatively between EOsAD and ADAD, which was unanticipated since the expression of pathology, as reflected in classical biomarkers, might be expected to be less in the EOsAD group, which had no family history of EOAD. Instead, we found that a small proportion even had *APP* and *PSEN* mutations ruled out by genetic testing. This outcome suggests that an ADAD-like phenotype without the accompanying genotype, occurs in EOsAD, a group which represents the majority of early onset cases, but which have received surprisingly little attention to date. Other potential emerging markers of neurodegeneration and inflammation, including CSF NfL, IgG, and IL-6, were reported in EOsAD. However, the relative paucity of additional biomarkers and low numbers of reported studies was notable, particularly for the relatively accessible serum/plasma sample type.

A β peptide is a hallmark of AD, whose complex lifecycle not only produces a plethora of chain

length variants such as A β_{38} , A β_{40} , A β_{42} , and A β_{43} [20], but also multimeric aggregate structures such as oligomers, fibrils, and plaque [21]. Of these, the 42-amino-acid peptide (A β_{42}) is the most studied variant due to its enhanced tendency to aggregate [22]. This process, known as the sink model, results in lower levels of soluble A β_{42} as plaque/insoluble levels increase [23]. Our meta-analysis results, which show a drop in CSF A β_{42} in all EOAD variants (Table 1), are consistent with this model. These observations are consistent with previous meta-analytic data on LOAD CSF A β_{42} [24]. Interestingly, the lowest pooled SMD (highest drop in A β_{42}) was for ADAD symptomatic mutation carriers (sMCs) and the EOsAD subset with no *APP*, *PSEN1*, or *PSEN2* mutations. The decreased level of CSF A β_{42} in ADAD aMCs was almost equal to EOsAD and suggested an EOsAD etiology similar to ADAD. There were few studies of other CSF A β peptides in the context of EOAD. However, a significant reduction in the level of A β_{43} in EOAD patients relative to controls has been reported [25].

Tau is one of the most abundant proteins in neurons with a role in stabilizing the structure of microtubules [26]. In AD, tau is hyperphosphorylated, dissociated from microtubules, and can be detected in CSF at an increased level relative to healthy controls [27]. Our meta-analyses showed a significant increase in both variants of tau (T-tau and P-tau) in all subtypes of EOAD. The pooled SMD of T-tau was significantly higher in ADAD sMCs as compared to EOsAD, likely reflecting a higher degree of neuronal damage in ADAD as compared to the EOsAD variant. The EOsAD subset with confirmed absence of *APP*, *PSEN1/PSEN2* mutations, had higher levels of both tau variants than the full EOsAD group. While the significance of this is not entirely clear, it suggests the possibility that the EOsAD group has mixed etiologies. Tau can be phosphorylated on various sites; however, most assays report phosphorylation of thr181 and ser199 specific epitopes because of the higher diagnostic accuracy of these epitopes to differentiate AD from healthy controls [28]. While our meta-analysis results confirm that these two specific epitopes discriminate between control and AD participants, neither the SMD nor the p-values were higher than for the general P-tau measure, suggesting little advantage in using these specific phospho-epitopes. In LOAD, several studies have reported other tau epitopes; thr231 was detected before the formation of paired helical filaments [28], whereas thr181 and ser199 were considered late events in AD [29].

According to some models, neurodegeneration is more closely linked to symptomatology as compared to A β pathology [30]. A handful of additional neurodegeneration and inflammation markers, including NfL, albumin, IL-6, and IgG that we identified as suitable for meta-analysis, were reported in only 2 to 4 studies each. Of these, CSF NfL was significantly upregulated in EOAD cases when compared to healthy controls. NfL is the light chain protein of neurofilament, which makes up neurofilament bundles with counterparts that determine the axonal caliber and conduction velocity [24]. NfL is an integral component of synapses and loss of synapses may account for the increasing level of NfL in blood and CSF [31, 32].

Furthermore, recent studies have established a correlation between blood and CSF NfL in late-onset AD [33], making it a potentially useful biomarker for neuronal apoptosis. Like tau, NfL is associated with the neuronal cytoskeleton, and interestingly, we found a similar increase in the pooled SMD of CSF NfL in EOsAD relative to controls (Supplementary Table 1). These findings relating to CSF levels of NfL in EOsAD suggest that additional studies in both CSF and serum/plasma are warranted, since an additional potential biomarker has been identified, even within this narrow range of replicated markers. There are several other well studied markers of neurodegeneration such as sTREM2, VILIP-1, and NfH, reported in LOAD [34, 35] which have not, as yet, been explored in EOsAD and ADAD.

A particularly notable feature of our EOAD meta-analysis was the lack of replicated data on blood biomarkers. Blood biomarker studies are more prevalent in LOAD [24, 36], whereas the EOAD biomarkers were assayed almost exclusively in CSF (Table 1). This may be a reflection of clinical practice, whereby CSF may be collected more often in EOAD cases, and less frequently in LOAD. Assay of CSF A β and tau for the diagnostic evaluation of dementia has recently been recommended by an international consensus group [37], but the invasive nature of a spinal tap makes it unlikely that it will be taken up widely [38]. For plasma A β ₄₂, a comprehensive meta-analysis was performed reporting a significant decrease in plasma A β ₄₂ in clinically diagnosed LOAD cases [36].

In contrast, our meta-analysis of ADAD (all MCs) shows a significant increase in plasma A β ₄₂, likely reflecting the higher peptide level in the plasma due to increased expression of A β ₄₂ or cleavage of APP at this earlier age group, in-line with the A β sink

model. Previous studies have reported that an increase in plasma level of A β ₄₂ correlates with the risk of AD development. With disease progression, the reduction of A β ₄₂ level over time has been recorded in plasma, reflecting deposition in plaque [39, 40]. In this younger cohort, clearance to plasma may be more efficient. There were insufficient studies on plasma A β ₄₂ in EOsAD, so a meta-analysis was not possible. Our meta-analysis on EOAD identified several studies reporting serum/plasma albumin and IgG in which a significant increase of IgG was observed (Table 1). These observations reflect findings in previous studies that show increased levels of both inflammation mediating proteins and pro-inflammatory IgG, which enhance the inflammatory activity in blood and represent a high-risk factor for AD [41]. In ADAD, serum NfL is emerging as a promising biomarker in the literature, but we could not perform a meta-analysis on NfL due to an insufficient number of studies. A longitudinal study, however, reported that an elevated level of NfL could differentiate mutation carriers from non-mutation carriers much earlier as compared with absolute NfL levels assayed in cross-sectional studies [18].

Interestingly, discovery-based approaches have the potential to quantify 100s–1000s of proteins, and while the majority are unlikely to be significantly changed in AD, the few that do change may provide not only critical diagnostic tools but also an insight into disease mechanisms. Plasma protein profiling of symptomatic and asymptomatic ADAD patients from the DIAN cohort reported significant differential expression of seven and sixteen proteins in aMC and sMC, respectively. These proteins included complement and inflammation family members which are known to exacerbate the pathogenic processes of AD [42]. Promisingly, an essential role for complement proteins in neuronal pruning, has recently been identified [43, 44].

The *APOE* ϵ 4 allele is the strongest genetic risk factor for LOAD [46]. In LOAD, *APOE* ϵ 4 is associated with increased A β levels and risk of A β accumulation at an earlier age, thereby accelerating the clinical onset of AD [45]. Furthermore, 50% of *APOE* ϵ 4 homozygous carriers have diagnosed AD by 70 years of age [46]. Less is known about *APOE* ϵ 4 allele carrier enrichment in EOAD. We identified significant enhancement in the level of the *APOE* ϵ 4 allele in EOsAD as compared to healthy controls: 43% and 30% respectively. By contrast, the ADAD group had a similar level of *APOE* ϵ 4 allele expression as compared to healthy controls; 28% and

29% respectively, indicating no significant difference (Table 1B). The level of reported *APOE* $\epsilon 4$ allele enrichment in LOAD as compared to controls was in the range of 13.7–40% [47, 48]. Our meta-analysis shows EOsAD *APOE* $\epsilon 4$ allele enrichment within the range of that reported for LOAD, and therefore enrichment of this allele does not account for the earlier age of onset of EOsAD relative to LOAD. Previous studies indicate a dose-dependent effect of *APOE* $\epsilon 4$ on CSF $A\beta_{42}$ level in LOAD [49]; however insufficient study numbers were available to explore an *APOE* $\epsilon 4$ dose effect for EOAD. EOsAD cohorts may represent an opportunity to explore novel genetic and lifestyle risk factors responsible for the earlier onset of symptoms in EOsAD.

Limitations

Some of the limitations of this study include the lack of longitudinal data, small sample size cohorts, and fewer studies which include analysis of plasma samples. For ADAD studies, the majority of the meta-analyses were based on two cohorts: DIAN and the Colombian Alzheimer's Prevention Initiative. Considerable heterogeneity was observed across studies, which was probably due to variation in study characteristics. We managed this in part by using a random effects model. Only 5 out of our 28 meta-analyses had sufficient study numbers to allow assessment of publication bias. The majority of EOsAD studies (16 out of 20) did not perform genetic analyses to eliminate *APP/PSEN* gene mutations and relied purely on the absence of family history. Interestingly, those 4 EOsAD studies in which *APP/PSEN* gene mutations were eliminated had lower levels of $A\beta_{42}$ and higher levels of all tau variants, suggesting an ADAD-like phenotype without the ADAD genotype, indicating that investigation of additional "non-classical" drivers of EOsAD is warranted. Furthermore, study numbers of particular biomarkers (i.e., CSF IL-6, IgG, and albumin) are low, so in some cases insignificant results could be a reflection of insufficient numbers of published studies, again reinforcing the need for additional biomarker work, particularly for genetically well-characterized EOsAD.

Conclusions

Our meta-analysis of established biomarkers has confirmed the quantitative difference between three variants of EOAD versus corresponding control groups. CSF NfL and IgG were statistically sig-

nificant and emerging as potential markers of neurodegeneration and inflammation to differentiate EOsAD from healthy individuals. EOsAD is a relatively neglected AD subtype of unknown etiology and therefore represents an opportunity to explore as yet unidentified genetic and/or lifestyle factors that contribute to earlier AD onset. To date, the range of biomarkers studied have been heavily focused on CSF, but for establishing promising blood tests, more research is needed before any of these could replace CSF or imaging tests. Generally, EOAD biomarker research is limited to the peptides considered to be hallmarks of AD pathophysiology whereas screening a broader spectrum of analytes (e.g., proteomic profiling, lipids, RNA, metabolites) together with longitudinal analysis is necessary to track the progress of disease pathology. Studies we included for meta-analysis were based on a dichotomization on age. However, studying age as a continuous variable in cohorts might provide insight into the trajectory of biomarker change with age and stage of disease. Similarly, a longitudinal study design would offer greater statistical power to identify biomarker changes over time. These are important considerations for the planning and design of new studies. Understanding pathophysiology and identification of additional preclinical biomarkers can be facilitated using discovery-based "omics" approaches.

ACKNOWLEDGMENTS

This work was supported by a National Health & Medical Research Council of Australia Program Grant (APP1054544) to PSS, UNSW Scientia Postgraduate Scholarship to GK, and Australian Research Council DECRA Postdoctoral Fellowship to NB. We thank the Rebecca L. Cooper Medical Research Foundation and the Sachdev Foundation for their ongoing research support.

We are thankful to Dr. Juan Marín, Dr. Erdinc Dursum, Dr. Steinunn Thordardottir, Prof. Anne Fagan, and Dr. Chengjie Xiong for providing data and information about the studies used in our meta-analyses.

Author's disclosures available online (<https://www.j-alz.com/manuscript-disclosures/20-0052r1>).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <https://dx.doi.org/10.3233/JAD-200052>.

REFERENCES

- [1] Isik AT (2010) Late onset Alzheimer's disease in older people. *Clin Interv Aging* **5**, 307-311.
- [2] Dai MH, Zheng H, Zeng LD, Zhang Y (2018) The genes associated with early-onset Alzheimer's disease. *Oncotarget* **9**, 15132-15143.
- [3] Bateman RJ, Aisen PS, De Strooper B, Fox NC, Lemere CA, Ringman JM, Salloway S, Sperling RA, Windisch M, Xiong C (2011) Autosomal-dominant Alzheimer's disease: A review and proposal for the prevention of Alzheimer's disease. *Alzheimers Res Ther* **3**, 1.
- [4] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health-care interventions: Explanation and elaboration. *BMJ* **339**, b2700.
- [5] Vandenberghe JP, Von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *Epidemiology* **18**, 805-835.
- [6] McDade E, Wang G, Gordon BA, Hassenstab J, Benzinger TL, Buckles V, Fagan AM, Holtzman DM, Cairns NJ, Goate AM, Marcus DS (2018) Longitudinal cognitive and biomarker changes in dominantly inherited Alzheimer disease. *Neurology* **91**, e1295-e1306.
- [7] Suárez-Calvet M, Caballero MÁ, Kleinberger G, Bateman RJ, Fagan AM, Morris JC, Levin J, Danek A, Ewers M, Haass C (2016) Early changes in CSF sTREM2 in dominantly inherited Alzheimer's disease occur after amyloid deposition and neuronal injury. *Sci Transl Med* **8**, 369ra178.
- [8] Fagan AM, Xiong C, Jasielec MS, Bateman RJ, Goate AM, Benzinger TL, Ghetti B, Martins RN, Masters CL, Mayeux R, Ringman JM (2014) Longitudinal change in CSF biomarkers in autosomal-dominant Alzheimer's disease. *Sci Transl Med* **6**, 226ra30.
- [9] Fleisher AS, Chen K, Quiroz YT, Jakimovich LJ, Gomez MG, Langois CM, Langbaum JB, Roontiva A, Thiyyagura P, Lee W, Ayutyanont N (2015) Associations between biomarkers and age in the presenilin 1 E280A autosomal dominant Alzheimer disease kindred: A cross-sectional study. *JAMA Neurol* **72**, 316-324.
- [10] Lim YY, Hassenstab J, Cruchaga C, Goate A, Fagan AM, Benzinger TL, Maruff P, Snyder PJ, Masters CL, Allegri R, Chhatwal J (2016) BDNF Val66Met moderates memory impairment, hippocampal function and tau in preclinical autosomal dominant Alzheimer's disease. *Brain* **139**, 2766-2777.
- [11] Thordardottir S, Ståhlbom AK, Almkvist O, Thonberg H, Eriksdotter M, Zetterberg H, Blennow K, Graff C (2017) The effects of different familial Alzheimer's disease mutations on APP processing *in vivo*. *Alzheimers Res Ther* **9**, 9.
- [12] Serghiou S, Goodman SN (2019) Random-effects meta-analysis: Summarizing evidence with caveats. *JAMA* **321**, 301-302.
- [13] Sedgwick P (2013) Meta-analyses: How to read a funnel plot. *BMJ* **346**, f1342.
- [14] Schoonenboom NS, Pijnenburg YA, Mulder C, Rosso SM, Van Elk EJ, Van Kamp GJ, Van Swieten JC, Scheltens P (2004) Amyloid β_{42} and phosphorylated tau in CSF as markers for early-onset Alzheimer disease. *Neurology* **62**, 1580-1584.
- [15] De Jong D, Jansen RW, Pijnenburg YA, Van Geel WJ, Borm GF, Kremer HP, Verbeek MM (2007) CSF neurofilament proteins in the differential diagnosis of dementia. *J Neurol Neurosurg* **78**, 936-938.
- [16] Hozo SP, Djulbegovic B, Hozo I (2005) Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* **5**, 13.
- [17] Schöll M, Wall A, Thordardottir S, Ferreira D, Bogdanovic N, Långström B, Almkvist O, Graff C, Nordberg A (2012) Low PiB PET retention in presence of pathologic CSF biomarkers in Arctic APP mutation carriers. *Neurology* **79**, 229-236.
- [18] Preische O, Schultz SA, Apel A, Kuhle J, Kaeser SA, Barro C, Gräber S, Kuder-Buletta E, LaFougere C, Laske C, Vögler J (2019) Serum neurofilament dynamics predicts neurodegeneration and clinical progression in presymptomatic Alzheimer's disease. *Nat Med* **25**, 277-283.
- [19] Weston PS, Poole T, Ryan NS, Nair A, Liang Y, Macpherson K, Druyeh R, Malone IB, Ahsan RL, Pemberton H, Klimova J (2017) Serum neurofilament light in familial Alzheimer disease. *Neurology* **89**, 2167- 2175.
- [20] Chow VW, Mattson MP, Wong PC, Gleichmann M (2010) An overview of APP processing enzymes and products. *Neuromolecular Med* **12**, 1-12.
- [21] Dahlgren KN, Manelli AM, Stine WB, Baker LK, Krafft GA, LaDu MJ (2002) Oligomeric and fibrillar species of amyloid β peptides differentially affect neuronal viability. *J Biol Chem* **277**, 32046-32053.
- [22] Jiang D, Rauda I, Han S, Chen S, Zhou F (2012) Aggregation pathways of the amyloid β_{42} peptide depend on its colloidal stability and ordered β -sheet stacking. *Langmuir* **28**, 12711-12721.
- [23] Schupf N, Tang MX, Fukuyama H, Manly J, Andrews H, Mehta P, Ravetch J, Mayeux R (2018) Peripheral A β subspecies as risk biomarkers of Alzheimer's disease. *Proc Natl Acad Sci U S A* **105**, 14052-14057.
- [24] Olsson B, Lautner R, Andreasson U, Öhrfelt A, Portelius E, Bjerke M, Hölttä M, Rosén C, Olsson C, Strobel G, Wu E (2016) CSF and blood biomarkers for the diagnosis of Alzheimer's disease: A systematic review and meta-analysis. *Lancet Neurol* **15**, 673-684.
- [25] Dorey A, Perret-Liaudet A, Tholance Y, Fourier A, Quadrio I (2015) Cerebrospinal fluid A β_{40} improves the interpretation of A β_{42} concentration for diagnosing Alzheimer's disease. *Front Neurol* **6**, 247.
- [26] Iqbal K, Alonso A, Gong C, Khatoon S, Kudo T, Singh T, Grundke-Iqbal I (1993) Molecular pathology of Alzheimer neurofibrillary degeneration. *Acta Neurobiol Exp* **53**, 325-335.
- [27] Mandelkow EM, Mandelkow E (1998) Tau in Alzheimer's disease. *Trends Cell Biol* **8**, 425-427.
- [28] Hampel H, Buerger K, Zinkowski R, Teipel SJ, Goernitz A, Andreasen N, Sjoegren M, DeBernardis J, Kerkman D, Ishiguro K, Ohno H (2004) Measurement of phosphorylated tau epitopes in the differential diagnosis of Alzheimer disease: A comparative cerebrospinal fluid study. *Arch Gen Psychiatry* **61**, 95-102.
- [29] Augustinack JC, Schneider A, Mandelkow EM, Hyman BT (2002) Specific tau phosphorylation sites correlate with severity of neuronal cytopathology in Alzheimer's disease. *Acta Neuropathol* **103**, 26-35.
- [30] Lashley T, Schott JM, Weston P, Murray CE, Wellington H, Keshavan A, Foti SC, Foiani M, Toombs J, Rohrer JD, Heslegrave A (2018) Molecular biomarkers of Alzheimer's

- disease: Progress and prospects. *Dis Model Mech* **11**, dmm031781.
- [31] Yuan A, Sershen H, Basavarajappa BS, Kumar A, Hashim A, Berg M, Lee JH, Sato Y, Rao MV, Mohan PS, Dyakin V (2015) Neurofilament subunits are integral components of synapses and modulate neurotransmission and behavior *in vivo*. *Mol Psychiatry* **20**, 986-994.
- [32] Bacioglu M, Maia LF, Preische O, Schelle J, Apel A, Kaeser SA, Schweighauser M, Eninger T, Lambert M, Pilotto A, Shimshek DR (2016) Neurofilament light chain in blood and CSF as marker of disease progression in mouse models and in neurodegenerative diseases. *Neuron* **91**, 56-66.
- [33] Zetterberg H (2016) Neurofilament light: A dynamic cross-disease fluid biomarker for neurodegeneration. *Neuron* **91**, 1-3.
- [34] Kester MI, Teunissen CE, Sutphen C, Herries EM, Ladenson JH, Xiong C, Scheltens P, Van Der Flier WM, Morris JC, Holtzman DM, Fagan AM (2015) Cerebrospinal fluid VILIP-1 and YKL-40, candidate biomarkers to diagnose, predict and monitor Alzheimer's disease in a memory clinic cohort. *Alzheimers Res Ther* **7**, 59.
- [35] Suárez-Calvet M, Kleinberger G, Caballero MÁ, Brendel M, Rominger A, Alcolea D, Fortea J, Lleó A, Blesa R, Gisbert JD, Sánchez-Valle R (2016) sTREM2 cerebrospinal fluid levels are a potential biomarker for microglia activity in early-stage Alzheimer's disease and associate with neuronal injury markers. *EMBO Mol Med* **8**, 466-476.
- [36] Song F, Poljak A, Valenzuela M, Mayeux R, Smythe GA, Sachdev PS (2011) Meta-analysis of plasma amyloid- β levels in Alzheimer's disease. *J Alzheimers Dis* **26**, 365-375.
- [37] Simonsen AH, Herukka SK, Andreasen N, Baldeiras I, Bjerke M, Blennow K, Engelborghs S, Frisoni GB, Gabryelewicz T, Galluzzi S, Handels R (2017) Recommendations for CSF AD biomarkers in the diagnostic evaluation of dementia. *Alzheimers Dement* **13**, 274-284.
- [38] Thambisetty M, Lovestone S (2010) Blood-based biomarkers of Alzheimer's disease: Challenging but feasible. *Biomark Med* **4**, 65-79.
- [39] Mayeux R, Honig LS, Tang MX, Manly J, Stern Y, Schupf N, Mehta PD (2003) Plasma A β 40 and A β 42 and Alzheimer's disease - Relation to age, mortality, and risk. *Neurology* **61**, 1185-1190.
- [40] Findeis MA (2007) The role of amyloid β peptide 42 in Alzheimer's disease. *Pharmacol Ther* **116**, 266-286.
- [41] Lundström SL, Yang H, Lyutvinskiy Y, Rutishauser D, Herukka SK, Soininen H, Zubarev RA (2014) Blood plasma IgG Fc glycans are significantly altered in Alzheimer's disease and progressive mild cognitive impairment. *J Alzheimers Dis* **38**, 567-579.
- [42] Muenchhoff J, Poljak A, Thalamuthu A, Gupta VB, Chatterjee P, Raftery M, Masters CL, Morris JC, Bateman RJ, Fagan AM, Martins RN (2016) Changes in the plasma proteome at asymptomatic and symptomatic stages of autosomal dominant Alzheimer's disease. *Sci Rep* **6**, 29078.
- [43] Hong S, Beja-Glasser VF, Nfonoyim BM, Frouin A, Li S, Ramakrishnan S, Merry KM, Shi Q, Rosenthal A, Barres BA, Lemere CA (2016) Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science* **52**, 712-716.
- [44] Chung WS, Verghese PB, Chakraborty C, Joung J, Hyman BT, Ulrich JD, Holtzman DM, Barres BA (2016) Novel allele-dependent role for APOE in controlling the rate of synapse pruning by astrocytes. *Proc Natl Acad Sci U S A* **113**, 10186-10191.
- [45] Huynh RA, Mohan C (2017) Alzheimer's disease: Biomarkers in the genome, blood, and cerebrospinal fluid. *Front Neurol* **8**, 102.
- [46] Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small G, Roses AD, Haines JL, Pericak-Vance MA (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921-923.
- [47] Liu CC, Kanekiyo T, Xu H, Bu G (2013) Apolipoprotein E and Alzheimer disease: Risk, mechanisms and therapy. *Nat Rev Neurol* **9**, 106-118.
- [48] Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, Van Duijn CM (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta-Analysis Consortium. *JAMA* **278**, 1349-1356.
- [49] Vemuri P, Wiste HJ, Weigand SD, Knopman DS, Shaw LM, Trojanowski JQ, Aisen PS, Weiner M, Petersen RC, Jack Jr CR (2010) Effect of apolipoprotein E on biomarkers of amyloid load and neuronal pathology in Alzheimer disease. *Ann Neurol* **67**, 308-316.
- [50] Ikeda M, Yonemura K, Kakuda S, Tashiro Y, Fujita Y, Takai E, Hashimoto Y, Makioka K, Furuta N, Ishiguro K, Maruki R, Yoshida J, Miyaguchi O, Tsukie T, Kuwano R, Yamazaki T, Yamaguchi H, Amari M, Takatama M, Harigaya Y, Okamoto K (2013) Cerebrospinal fluid levels of phosphorylated tau and A β 1-38/A β 1-40/A β 1-42 in Alzheimer's disease with PS1 mutations. *Amyloid* **20**, 107-112.

Extending the Depth of Human Plasma Proteome Coverage Using Simple Fractionation Techniques

Gurjeet Kaur, Anne Poljak,* Syed Azmal Ali, Ling Zhong, Mark J. Raftery, and Perminder Sachdev

Cite This: *J. Proteome Res.* 2021, 20, 1261–1279

Read Online

ACCESS |

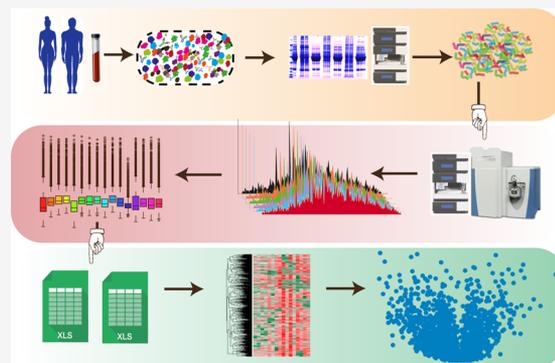
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Human plasma is one of the most widely used tissues in clinical analysis, and plasma-based biomarkers are used for monitoring patient health status and/or response to medical treatment to avoid unnecessary invasive biopsy. Data-driven plasma proteomics has suffered from a lack of throughput and detection sensitivity, largely due to the complexity of the plasma proteome and in particular the enormous quantitative dynamic range, estimated to be between 9 and 13 orders of magnitude between the lowest and the highest abundance protein. A major challenge is to identify workflows that can achieve depth of plasma proteome coverage while minimizing the complexity of the sample workup and maximizing the sample throughput. In this study, we have performed intensive depletion of high-abundant plasma proteins or enrichment of low-abundant proteins using the Agilent multiple affinity removal liquid chromatography (LC) column—Human 6 (Hu6), the Agilent multiple affinity removal LC column—Human 14 (Hu14), and ProteoMiner followed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS PAGE) and C18 prefractionation techniques. We compared the performance of each of these fractionation approaches to identify the method that satisfies requirements for analysis of clinical samples and to include good plasma proteome coverage in combination with reasonable sample output. In this study, we report that one-dimensional (1D) gel-based prefractionation allows parallel sample processing and no loss of proteome coverage, compared with serial chromatographic separation, and significantly accelerates analysis time, particularly important for large clinical projects. Furthermore, we show that a variety of methodologies can achieve similarly high plasma proteome coverage, allowing flexibility in method selection based on project-specific needs. These considerations are important in the effort to accelerate plasma proteomics research so as to provide efficient, reliable, and accurate diagnoses, population-based health screening, clinical research studies, and other clinical work.

KEYWORDS: plasma, proteome, label-free quantitation, biomarkers, prefractionation techniques, mass spectrometry, chromatography, electrophoresis



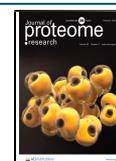
INTRODUCTION

Human blood, plasma, and serum are clinical samples commonly used for diagnostic analyses as they are minimally invasive sample types, already in routine clinical use, and are in direct contact with damaged tissues and therefore frequently reflect downstream effects of disease processes.^{1–3} These sample types contain a diversity of analytes, such as metabolites, lipids, and proteins/peptides, of which proteins/peptides are commonly used as biomarkers, since arguably, they represent the full diversity of cellular machinery, are the drivers of all biochemical pathways, and are the functional counterpart to the genome. Consequently, they are the most direct and specific targets of cellular and physiological homeostases or change. Immune assays, such as enzyme-linked immunosorbent assay (ELISA), are commonly used for quantitative analysis of individual proteins for disease diagnosis. However, they are inherently targeted approaches, which are useful for aiding diagnosis of diseases where the

pathophysiology is understood, but are of limited use for discovery-based/data-driven investigations, even when used in multiplexed assays.⁴ Furthermore, antibody-based approaches frequently lack specificity for proteins and/or their isoforms. Consequently, mass spectrometry (MS)-based technology is not only the golden standard for targeted quantification but also the only approach that allows unbiased, data-driven investigations. In clinical research, mass spectrometry is, therefore, a method of choice for both absolute and relative quantification of proteomic expression changes, exploration of disease mechanisms, and biomarker discovery in plasma.⁵

Received: August 29, 2020

Published: January 20, 2021



However, the extended dynamic range of protein abundance in plasma, reported to range between 9 and 13 orders of magnitude, makes MS-based plasma proteomics extremely challenging.^{6,7}

Over the last 3 decades, proteomics technologies have evolved and improved, so that identification of hundreds to thousands of proteins in various tissue samples is now routine. Nonetheless, relatively few biomarkers identified by proteomics techniques have progressed to use in clinical practice.^{8–10} This is in part due to the lack of robust, reproducible, and high-throughput proteomics workflows to determine and validate potential biomarkers in large cohorts.

Some low-abundance proteins (LAPs) may be transient in plasma as a result of variations in the disease state, temporary tissue damage, specific diet, and other environmental factors and may remain undetected. Furthermore, the overwhelming presence of high-abundance proteins (HAPs), which represent >99% of the total plasma protein mass,¹¹ can obscure lower abundance components from detection by mass spectrometry. Plasma prefractionation methods therefore play a crucial role in overcoming the complexity of plasma samples, providing the opportunity to identify low-abundance proteins.¹² More importantly, tissue-specific proteins, which are more likely to be clinically relevant, are likely to be of medium to low abundance and therefore are more likely to be identified in fractionated plasma. Various standard plasma prefractionation approaches have been reported in the literature, including affinity enrichment, immunodepletion, and a variety of chromatographic and electrophoretic fractionation techniques. In general, the greater the plasma proteome coverage, the more complex and time-consuming the prefractionation workflow,^{13–28} making most methods impractical for large clinical studies.

The objective of this study was to identify a fractionation strategy that would provide good plasma proteome coverage and identify tissue-specific proteins while being compatible with the higher sample throughput required of many clinical studies. Using a two-step approach, we (1) compared high-abundance protein depletion (immunoaffinity columns; Hu6, Hu14) and low-abundance protein enrichment (ProteoMiner combinatorial peptide ligands), followed by (2) further fractionation of the low-abundance proteins using chromatographic and electrophoretic approaches (C18 column and one-dimensional (1D) sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS PAGE), respectively).

■ EXPERIMENTAL SECTION

Experimental Procedures

To optimize the plasma fractionation methodology, whole blood was collected by venipuncture from two healthy volunteers, one male and one female, within the 35–55 year age range. A single pooled sample was used to compare workflows and to avoid confounding the method comparisons by biological variation. Blood was collected into ethylenediaminetetraacetic acid (EDTA)-containing tubes and centrifuged (2000g, 20 min, 4 °C), and the plasma was transferred and pooled into a clean 15 mL polypropylene tube. To minimize freeze–thaw cycles, plasma aliquots were prepared (250–500 μ L) and stored at -80 °C until required. The UNSW Human Research Ethics Committee approved a protocol for blood collection. The blood samples used here

were collected from volunteers, and both subjects gave written informed consent.

Depletion of High-Abundant Proteins Using Immunoaffinity-Based Columns: Human 6 (HU6) and Human 14 (HU14) Columns

We started our experiments with a total volume of 900 μ L of plasma, which was divided equally into three vials (each 300 μ L), one for each of the three prefractionation techniques (Hu6, Hu14, and ProteoMiner). Depletion of 300 μ L of crude plasma using antibody-based immunoaffinity columns HU6 (4.6×50 mm², Agilent, CA) and HU14 (4.6×100 mm², Agilent, CA) was achieved as follows: *Plasma preparation for the HU6 column:* 300 μ L of plasma was diluted with 1200 μ L of buffer A (1:5 dilution, as recommended by Agilent Technologies), and each injection was 100 μ L on the HU6 column (composed of 20 μ L of plasma + 80 μ L of buffer A). This step eliminates approximately 85–90% of the total plasma protein by binding the six high-abundance plasma proteins (albumin, IgG, IgA, transferrin, haptoglobin, and antitrypsin). *Plasma preparation for the Hu14 column:* 300 μ L of plasma was diluted with 900 μ L of buffer A (1:4 dilution, as recommended by Agilent Technologies), and each injection was 120 μ L on the HU14 column (composed of 30 μ L of plasma + 90 μ L of buffer A). This step eliminates approximately 94% of the total plasma protein and contains antibodies to the following 14 high-abundance plasma proteins (albumin, IgG, antitrypsin, IgA, transferrin, haptoglobin, fibrinogen, α -2-macroglobulin, α -1-acid glycoprotein, IgM, apolipoprotein AI, apolipoprotein AII, complement C3, and transthyretin). Plasma was diluted and filtered using spin filters (Corning Costar Spin-X centrifuge tube filter, 0.45 μ m cellulose acetate, Merck, Germany) before chromatography to remove any particulates. Chromatography and fraction collection were performed on an Agilent 1290 ultrahigh-performance liquid chromatography (UHPLC) system (Agilent, Santa Clara, CA). Binding buffer (A) and elution buffer (B) were commercial proprietary products, and the liquid chromatography (LC) methods used were based on the manufacturer's instructions (Agilent, Santa Clara, CA). Highly abundant proteins were retained on the columns while buffer A was in use, and the flowthrough proteins represented the low-abundance protein fraction. The high-abundance proteins were then eluted with buffer B. Only the low-abundance protein fractions were processed for further fractionation and LC-tandem mass spectrometry (MS/MS) analysis.

Enrichment of Low-Abundant Proteins Using Bead-Based Depletion

We sought to compare high-abundance protein depletion against a low-abundance protein enrichment approach using beads bound to complex mixtures of combinatorial peptide ligands (ProteoMiner beads; Bio-Rad, Berkeley, CA). The ProteoMiner bead preparation consisted of 500 μ L of bead slurry (20% beads in aqueous ethanol, 20% v/v), resulting in 100 μ L of settled bead volume. Plasma (300 μ L) was used for depletion, with the manufacturer manual recommending 200 μ L of plasma sample with a protein concentration of 70 mg/mL. The beads and plasma mixture was incubated for 2 h at ambient temperature on a rotation platform. The sample was then centrifuged (1000g, 1 min, ambient temperature), and the flowthrough fraction containing the unbound proteins was collected for further analysis. Based on the manufacturer's protocol, we performed three washing steps using a wash

buffer, discarding the solution from each wash step. Therefore, the unbound fraction does not include any proteins that may have been present in the wash steps. This approach reduces the dynamic range of plasma proteins since low- and high-abundance proteins have a similar chance of being bound to their respective ligands, which are distributed in similar proportions. Excess high-abundance proteins are washed away once they reach their ligand saturation point, whereas low-abundance proteins may never saturate their ligand and are enriched within the sample. Both bound and unbound (flowthrough) samples were further fractionated (C18 or SDS PAGE) and analyzed by LC-MS/MS.

Depleted plasma proteins from the above approaches were concentrated and buffer-exchanged to 50 mM ammonium bicarbonate back to the original volume (300 μ L) using Amicon 3 kDa concentrators (Millipore, Billerica, MA). Protein concentrations of depleted plasma were determined by use of a nanoliter-scale spectrophotometer (DeNovix DS-11+, POCD, Sydney, Australia) used in the microvolume mode with 2 μ L of each 3 kDa concentrated sample, 280 nm wavelength, and using the extinction coefficient of bovine serum albumin (BSA) ($E_{1\%}^{1\text{cm}} = 6.67$). Following absorbance reading, the samples were retrieved from the DeNovix sample surface to minimize total protein loss. The step was repeated three times, and mean values were used as the final sample concentration. To compare the protein profiles across all methods, 50 μ g of concentrated plasma from all methods was used as the starting point for further fractionation approaches. All subsequent fractionations were each performed in duplicate, providing two technical replicates for each fractionation approach. In this way, the depleted plasma samples from each of the Hu6, Hu14, ProteoMiner bound (PB), and ProteoMiner unbound (PUB) steps (outlined above) were fractionated in duplicate by each of the 1D-SDS PAGE and C18 methods (explained below). Therefore, each step generated 24 fractions \times 2, and all were analyzed separately by LC-MS/MS using all eight prefractionation methods, each using an identical pooled aliquot of depleted plasma.

Electrophoretic Fractionation of Low-Abundance Proteins Using 1D-SDS PAGE

Equal amounts of protein samples from crude and depleted plasma (50 μ g), obtained from low- and high-abundant fractions of HU6, HU14, ProteoMiner bound (PB), and ProteoMiner unbound (PUB) methods, were prepared in duplicate and diluted in 5 μ L of LDS sample buffer (4 \times NuPAGE, Invitrogen), 2 μ L of reducing agent (10 \times NuPAGE, Invitrogen), and deionized water to make a final volume of 20 μ L. Samples were heated for 10 min at 70 $^{\circ}$ C, and proteins were separated on NuPAGE 4–12% Bis–Tris midi gels (Invitrogen) using 1 \times SDS PAGE running buffer and the following electrophoretic parameters: voltage 200 V, current 200I for 54 min (NuPAGE, Invitrogen). Gels were then stained²⁹ (colloidal coomassie G250). After destaining, the separated protein lanes were cut into 24 equal-sized gel bands using a gel cutter tool with a 24-lane blade (Gel Company Inc., CA) and placed into 24 vials for in-gel trypsin digestion.

In-Gel Trypsin Digestion

In-gel digestion was performed by adaptation of previously published approaches.³⁰ Prior to performing in-gel digestion, excised gel bands (duplicate gel lanes/depletion method) were destained by incubating in 100 mM ammonium bicarbonate/

acetonitrile (1:1, 200 μ L) overnight at room temperature. Proteins were then reduced in an aqueous solution of 100 mM ammonium bicarbonate containing 100 mM dithiothreitol (DTT) (100 μ L) and incubated at 37 $^{\circ}$ C for 1 h. Proteins were alkylated using 20 μ L of 200 mM iodoacetamide (IAA) and incubated at 37 $^{\circ}$ C for 30 min. The reduction/alkylation solution was decanted, and the gel bits were further washed using 100% acetonitrile to remove DTT and IAA reagents. Trypsin digestion (Promega, sequence-grade modified trypsin) was performed (10 ng/2 μ L) in 100 μ L of 10 mM ammonium bicarbonate at 37 $^{\circ}$ C for \sim 18 h. Formic acid (0.1%, 20 μ L) was added and incubated at ambient temperature (10 min) followed by addition of 100% acetonitrile (50 μ L). The extracted peptide solution was transferred to LC vials (Agilent Technologies) and dried using a centrifugal evaporator (SpeedVac, Thermo Fisher Scientific). Peptides were redissolved in 5 μ L of 0.1% heptafluorobutyric acid containing 1% formic acid and were then ready for LC-MS/MS analysis.

In-Solution Protein Digestion

For chromatographic peptide fractionation, proteins were digested in situ using 50 μ g of the initial protein sample from all depletion methods in duplicate. Proteins were reduced by adding 2 μ L of tris-(2-carboxyethyl)phosphine (TCEP) and incubated for 60 min at 60 $^{\circ}$ C. Proteins were further alkylated using 1 μ L of 200 mM iodoacetamide (37 mg IAA/mL), vortexed, spun briefly, and incubated at ambient temperature for 10 min. The sample pH was checked and adjusted to pH 7–9 if necessary; then, trypsin was added (4.4 μ g/10 μ L) (Promega, sequence-grade modified trypsin) and incubated at 37 $^{\circ}$ C for \sim 16 h. The extracted peptide solution was transferred to LC vials (Agilent Technologies) and dried using a centrifugal evaporator (SpeedVac, Thermo Fisher Scientific). Peptides were redissolved in 5 μ L of 0.1% heptafluorobutyric acid containing 1% formic acid and were then ready for LC-MS/MS analysis.³¹

Chromatographic Fractionation of Low-Abundance Proteins Using a C18 Column (ZORBAX Extended-C18)

Chromatographic separation of tryptic peptides was achieved using a binary gradient; buffer A consisted of 20 mM ammonium formate containing 2% acetonitrile (pH 10) and buffer B consisted of 20 mM ammonium formate containing 90% acetonitrile (pH 10). The 64.5 min gradient was started with a 5 min isocratic step of 100% buffer A, followed by several gradient steps: 0–15% buffer B in 8 min; 15–28.5% buffer B in 33 min; 28.5–34% buffer B in 5.5 min; 34–60% buffer B in 13 min. Peptide fractions were collected at 0.6 min time intervals, resulting in a total of 84 fractions. These were then concatenated by pooling equally spaced early, mid-, and late fractions, and the pattern was repeated to a final number of 24 peptide fractions. The concatenation resulted in fractions consisting of peptides across the full chromatographic timespan. The fractions were dried by vacuum centrifugation (SpeedVac, Thermo Fisher Scientific), resuspended in 1% formic acid containing 0.1% heptafluorobutyric acid (HFBA), and were ready for LC-MS/MS.

Mass Spectrometry

Technical duplicates of each sample were analyzed using a Q-Exactive Plus mass spectrometer (Thermo Electron, Bremen, Germany) using an adaptation of a previously published work.³² Peptide separation was carried out using a nano-LC, Dionex UltiMate 3000 high-performance liquid chromatog-

raphy (HPLC) system (Thermo Scientific, Waltham), equipped with an autosampler (Dionex, Amsterdam, the Netherlands). Peptides ($\sim 2 \mu\text{g}$ on-column) were initially captured onto a C18 cartridge (Acclaim PepMap 100, 5 μm 100 Å, Thermo Scientific Dionex, Waltham), switching to a capillary column (25 cm length, 350 μm o.d., 75 μm i.d.) containing reverse-phase packing (C18, Reprosil-Pur, 1.9 μm , 200 Å, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany), placed within a column heater (45 °C, Sonation GmbH, Germany). Peptide elution involved a 60 min run time and a binary gradient of 0–45% buffer B at 200 nL/min. The binary buffers consisted of buffer A ($\text{H}_2\text{O}/\text{CH}_3\text{CN}$ of 98:2 containing 0.1% formic acid) and buffer B ($\text{H}_2\text{O}/\text{CH}_3\text{CN}$ of 20:80 containing 0.1% formic acid). Mass spectrometer settings were as follows: ion spray voltage 2000 V, capillary temperature 275–300 °C, positive ion mode. The method is a shotgun sequencing approach, using data-dependent acquisition (DDA), with a survey scan acquired (m/z 375–1750) and up to 10 multiply charged ions (charge state $\geq 2^+$) isolated for MSMS fragmentation (intensity threshold of 8.0×10^4 , with nitrogen as the higher-energy collisional dissociation (HCD) gas). The auto MS/MS acquisition consisted of 10 scans per duty cycle, an MS resolution of 30 000, and an MSMS resolution of 70 000.

Computational Analysis

Initially, the data processing of the raw files was performed on MaxQuant (version 1.6.10.43)³³ using the Andromeda search engine.³⁴ For profiling and quantitation, we used the filter of ≥ 2 unique peptides per protein. The UniProt *Homo sapiens* database (Proteome ID UP000005640, SwissProt, and TrEMBL) was used, in combination with reverse decoy protein sequences (used for protein identification false discovery rate (FDR) estimation) and frequent contaminant sequences. The default values of precursor mass tolerance (± 0.07 Da) for the first search and (± 0.006 Da) for the main search were applied. Mass tolerance for matching peaks to theoretical MSMS ion series was 5 ppm. The protein identification FDR was set to $< 1\%$ to ensure only high-confidence identifications for the peptide to spectra match, protein, and site decoy fractions. Enzyme specificity was set to trypsin, with a maximum of two missed cleavages. Searches included variable modifications of protein N-terminal acetylation, methionine oxidation, pyroGlu, and pyroGln and fixed modification of carbamidomethylation of cysteines. The search tolerance used for both peptide precursor match and fragment ions was 20 ppm. The search tolerance used for both peptide precursor match and fragment ions was 20 ppm, and only those peptides of more than six amino acid residues in length were used. Peptide spectral matches (PSMs), peptides, and protein were all selected based on the false discovery rate (FDR) of less than 0.01 based on a reversed-sequence decoy database. Furthermore, following filtering with these criteria, only peptides with ≥ 2 PSM were selected, and only proteins identified with a minimum of two unique peptides were included. To validate and transfer identifications across adjacent fractions and across technical replicates, the “match between runs” (MBR) option in MaxQuant was enabled, with a retention time window of 0.7 min and an alignment time window of 20 min. The full parameter set for the MaxQuant analysis is shown in Table S6.

All of the commonly identified proteins in all methods were quantified for peak area abundance within the MaxQuant

environment using the MaxLFQ (label-free quantification) algorithm.³⁵ The MaxLFQ algorithm combines the measured intensities from all matched peptides combined into proteins for final abundance data.³⁶ Finally, the MaxLFQ values obtained from MaxQuant for each protein were statically analyzed in the Perseus environment. This workflow of Andromeda to MBR to MaxLFQ followed with Perseus is frequently used in studies using data-dependent quantitative proteomics.³⁷

All of the raw files were also processed through the Trans-Proteomic Pipeline (TPP) version 5.1.0 released on 2017-11-03 using an adaptation of previously published works.^{38,39} This software takes into account additional factors, including the number of sibling experiments identifying the same peptide ions, the number of replicate ion identifications, sibling ions, and sibling modification states. TPP uses the iProphet algorithm to determine the false discovery rate (FDR) for protein identification. To perform the TPP analysis, the LC-MS/MS raw files were first converted to open format mzML files using the MSconvertGUI with the default parameters. The Uniprot database and sequence input were similar to the MaxQuant search. The 384 LC-MS/MS raw files were processed using three search engines: X! Tandem (with the k -score plug-in),⁴⁰ Mascot, and Comet, with all parameters kept similar. The PeptideProphet and ProteinProphet algorithms were employed to compute the probability scores for peptide and protein identification. The accurate mass model in PeptideProphet was used for high-confidence peptide identifications⁴¹ as well as a requirement that each protein be identified with a minimum of two peptides, each with a peptide probability score $> 99\%$. All three search engine results were merged and validated using iProphet.⁴² An iProphet probability of > 0.99 was used as the cutoff for final protein identification.⁴³ The total sum mapping resulted in the identification of 10 million spectra, but we used only the common peptides representing only the highest confidence data and without a search engine bias.^{40–44} Prior to further data processing of our confidently identified protein lists, contaminant proteins and a few remnant isoforms (mainly resulting from unreviewed UniProt TrEMBL data) were removed manually. The full list of manually removed proteins is shown in Tables S1–S4.

Density and Multiscatter Plots

All analyses were performed using a combination of MS Excel (<https://www.microsoft.com/en-us/microsoft-365/excel>), R-language (<https://www.r-project.org/about.html>) and Perseus (<https://maxquant.net/perseus/>) platforms. For density scatter plots and multiscatter plot gplots and ggplot2, packages in R were employed. Correlation was assessed for replicate LFQ intensities using the Pearson correlation coefficient scatter plots prepared by combining all of the experiments and calculating the Pearson correlation of the expression data across methods.

Hierarchical Clustering Analysis (HCA)

Pairwise comparison of sample distance and similarity analysis was performed using hierarchical clustering analysis (HCA). The k means clustering method was used for all protein LFQ intensities and was performed using the Heatmap package in R. The “one minus Pearson correlation coefficient” with the average linkage method was employed. Box plots were prepared using the ggplot2 package in R.

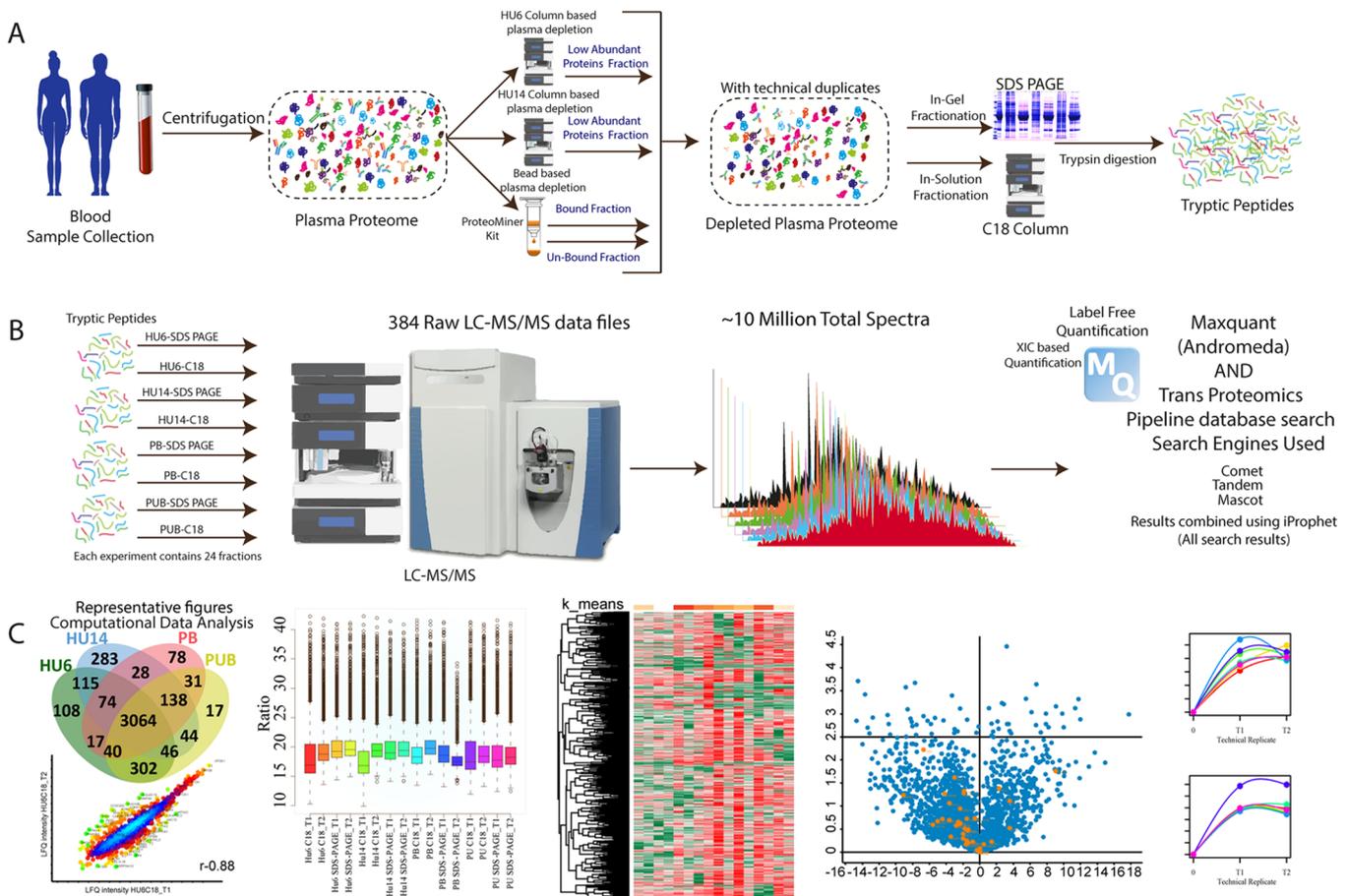


Figure 1. Schematic representation: workflow used for method optimization of plasma fractionation and proteomics analysis. (A) Fractionation approaches, (B) LC-MS/MS analysis and data processing. (C) Computational analysis.

Volcano Plot and Principal Component Analysis (PCA)

To compare depletion and fractionation approaches with the aim of identifying any significant protein fold change across methods, we used volcano plots and principal component analysis (PCA). Two-sided Student's *t*-test for *p*-value calculation was used for the volcano plot. The test parameter includes the minimum number of nine valid values in the total mode (all groups per row). Correction for multiple testing was performed using the Benjamini–Hochberg correction, with an FDR of less than 0.05 accepted as significant. We used the KEGG database for pathway annotation of the whole plasma proteome, using all 28 pairwise method comparisons, encompassing all 3064 proteins, which overlapped across methods. Significant differences in protein recovery/identification across methods were accepted with a $-\text{Log} P$ value >2.5 , and these were represented on volcano plots by differential coloring (yellow dots). The PCA analysis (plotted in the Perseus environment) was used to determine whether the separation methods and technical replicates would be differentiated globally.

Human Protein Atlas Mapping

We used data in the Human Protein Atlas database (<https://www.proteinatlas.org/>) together with the 3064 confidently identified proteins in our plasma proteome, which also overlapped across methods, to determine the tissue specificity of each protein. Confidently identified proteins were manually curated against the Protein Atlas lists, identifying the tissue

expression level across the 51 tissues and/or tissue fractions reported in The Human Protein Atlas. This was done to establish the level of tissue specificity of each of the plasma proteins we identified. The data was collated to represent proteins specific to each of the 11 human organ systems (digestive, circulatory, integumentary, nervous, urinary, skeletal, lymphatic, respiratory, muscular, endocrine, and reproductive).

RESULTS

Comparative Proteomic Data Provided by Plasma Depletion and Separation Workflow

The two-step plasma fractionation approaches that we implemented in this study (Figure 1) resulted in eight sample types (Figure 1B), each with 24 fractions and 2 technical replicates, which were analyzed by LC-MS/MS, and resulted in 384 raw files and a total of 2 890 105 spectra (Table 1), of which approximately 25% were assigned to peptides following processing with search engines. All raw data files have been uploaded to ProteomeXchange (<http://www.proteomexchange.org/>) with the identifier PXD02246. Comparison of the high-abundance protein fraction and the depleted plasma SDS PAGE profiles showed that substantial amounts of the higher-abundance proteins were extracted into the Hu6, Hu14, and PUB fractions, and the low-abundant protein fractions showed a much more complex banding pattern, in comparison to undepleted plasma (Figure S1). Using bioinformatics computational analysis in MaxQuant

Table 1. MaxQuant Outcomes^a

workflow	peptide fractionation approach	number of fractions	technical replicate	total number of queries spectra	assigned spectra	total peptides	total proteins per fractionation approach	total proteins	common proteins
HU6 (high-abundance protein removal)	SDS	24 fractions	T1	279 069	90 752	13 132	4209	3766	3064
			T2	282 075	82 494	12 355	4182		
			T1T2	561 144	173 766	15 080	4219		
	C18	24 fractions	T1	111 445	27 656	8238	3696		
			T2	148 551	31 216	8099	3652		
			T1T2	259 996	59 226	10 950	3680		
HU14 (high-abundance protein removal)	SDS	24 fractions	T1	217 542	54 719	11 536	4186	3792	
			T2	184 648	57 803	12 139	4125		
			T1T2	391 084	118 476	14 971	4190		
	C18	24 fractions	T1	103 220	26 985	8068	3819		
			T2	107 281	22 359	6241	3916		
			T1T2	210 501	49 816	9944	4010		
ProteoMiner bound (low-abundance protein enrichment)	SDS	24 fractions	T1	233 641	73 377	9949	2813	3470	
			T2	217 027	61 341	9164	2974		
			T1T2	450 668	134 718	11 239	3254		
	C18	24 fractions	T1	94 606	19 451	5567	3406		
			T2	90 944	23 686	7064	3812		
			T1T2	185 550	43 645	8663	3825		
ProteoMiner unbound (low-abundance protein enrichment)	SDS	24 fractions	T1	177 590	45 221	7774	4191	3682	
			T2	204 826	49 603	7858	3840		
			T1T2	382 416	95 521	9867	4198		
	C18	24 fractions	T1	217 538	47 400	6507	4009		
			T2	220 102	51 204	7820	4191		
			T1T2	437 640	99 296	9523	4211		

^aEquivalent amounts of total plasma low-abundance proteins (50 μ g) were prepared by three different methods of eliminating high-abundance protein components (HU6, HU14, and ProteoMiner) and two fractionation approaches (SDS and C18). Two technical replicates were acquired per sample (T1 and T2, with T1T2 representing a database search with combined technical replicates). Summarized data on numbers of spectra, peptides, and proteins acquired and identified are shown, including the total number of query spectra (all LC-MS/MS spectra acquired, including background and peptide signals), assigned spectra (these are the spectra for which a protein database match was identified), and identified peptides (these are the unique plus razor peptides identified). They are fewer than the assigned spectra because frequently multiple spectra for the same peptide are acquired due to the chromatographic peak width often exceeding the delay time set before the same mass precursor is again subject to LC-MS/MS, total proteins (the proteins identified for each fractionation approach and each workflow), and the common proteins (proteins that overlap across all workflows).

search engine Andromeda (Tables S1–S4) and the Trans Proteomics Pipeline (TPP) search engines Mascot, Comet, and XTandem (Table S5), we found that the commonly mapped transitions from all four search engines (Andromeda, Mascot, Comet, XTandem) resulted in the identification of 3597 high-confidence proteins (Table 1). A schematic representation of the workflows used for method development is shown in Figure 1, and results of the total protein numbers are summarized in Table 1 (Tables S1–S4 show the protein lists per specific depletion strategy).

The highest number of total proteins (4219) was identified with HU6 depletion SDS PAGE fractionation, and a similar number of proteins (4190) was identified with the HU14 depletion SDS PAGE fractionation approach (Table 1). The lowest number of proteins (3254) was identified in the PB low-abundance protein enrichment SDS PAGE fractionation method (Table 1). Of the ~4385 total proteins confidently identified, the majority are common to all methods, i.e., 3064, representing ~70% overlap across methods (Figure 2A), while 2664 and 2705 were in common across SDS PAGE and C18 fractionation techniques, respectively (Figure 2B,C). The detailed information for all of the identified LFQ intensities of 4385 proteins that were extracted and used for downstream analysis is shown in Tables S1–S4, representing the data for each depletion strategy, separately.

Technical Performance across Workflows

Technical replicates were highly correlated across all eight methods, with an average Pearson correlation coefficient of $r = 0.90$, and density scatter plots of technical replicates are shown in Figure 3. While the within-method technical replicate correlations are high, the cross-method correlations are lower. The detailed analysis is provided in the multiscatter correlation plot (Figure S2).

Box plots showing the distribution and variations of the LFQ intensities in all 16 analyses (8 methods, each with 2 technical replicates) are shown in Figure 4A. The line within the box denotes the median value, and the upper and lower ranges of the box indicate the 5 and 95 percentiles of the intensities, respectively. The dots above the box plots are outliers. Hierarchical clustering analysis (HCA) of the 16 analyses resulted in the formation of the eight clusters, with associated technical replicates grouped most closely, as shown in Figure 4B. Dendrogram-based hierarchical relationship (Figure 4B) and similarity matrix analyses (Figure 4C) show the close association of protein expression data between the following method pairs: HU6-SDS PAGE and HU14-SDS PAGE, HU6-SDS PAGE and PB-SDS PAGE, HU6-SDS PAGE and PUB-SDS PAGE, HU14-SDS PAGE and PB-SDS PAGE, and PB-SDS PAGE and PB C18.

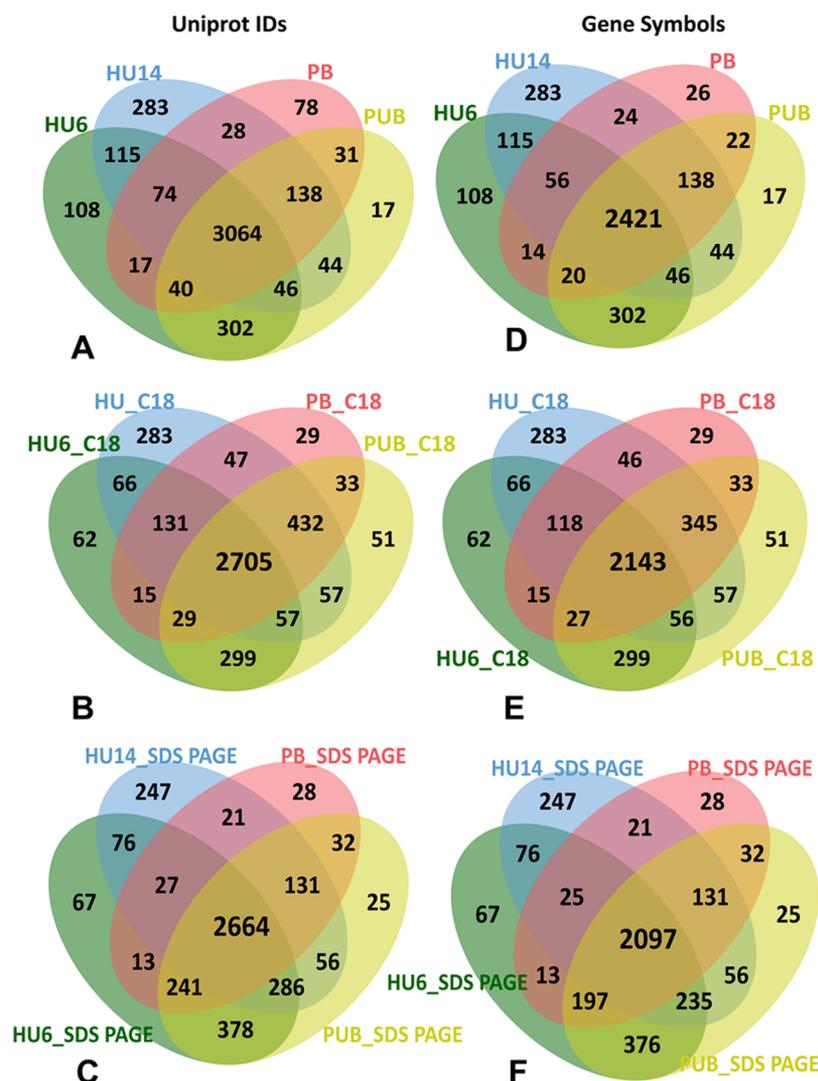


Figure 2. (A) Venn diagram showing the protein overlap across all depletion methods together (HU6, HU14, PB, and PUB) for all proteins identified with unique Uniprot accession numbers. (B) Method comparison of proteins with unique Uniprot accession numbers using the C18 fractionation approach. (C) Method comparison of proteins with unique Uniprot accession numbers using the SDS PAGE fractionation approach. (D) Venn diagram showing protein overlap across all depletion methods together (HU6, HU14, PB, and PUB) for proteins with both unique Uniprot accession numbers and gene symbols. (E) Comparison of proteins identified with both unique Uniprot accession numbers and gene symbols using the C18 fractionation approach. (F) Comparison of proteins identified with both unique Uniprot accession numbers and gene symbols using SDS PAGE fractionation.

Comparison of Different Prefractionation and Separation Techniques

We then compared the LFQ intensities merging the technical replicates to compare the C18 and SDS PAGE separation techniques (Figure 5). In the cases of the Hu6 and Hu14 depletion approaches, the profile of relative peak intensities (LFQ intensity) is shifted slightly to the right for the SDS PAGE relative to the C18 fractionation methods, indicating a greater peak intensity in the SDS PAGE methods. This observation is also recapitulated in Table 1, which shows higher numbers of spectra, peptides, and proteins for most SDS PAGE methods compared with C18 methods. Almost equal LFQ intensities were observed in the PB and PUB fractions, indicating that a substantial number of proteins was present in both fractions. The density overlay also shows similar signal intensities of proteome, indicating that both the number and intensity of proteins substantially overlap in the PB and PUB fractions.

Principal Component Analysis (PCA)

PCA was performed to explore the degree of differentiation or similarity between technical replicates and across methods. The first and second principal components explained 47.8 and 12.9% of data variability, respectively (Figure 6A). The PCA results show that the technical replicates cluster together as might be expected, and all methods are discriminated from one another (Figure 6A). However, closer positioning of specific methods on the PCA plot is observed, particularly in the principal component 1 (PC1) dimension, for the following method pairs: HU6-SDS PAGE and HU14-SDS PAGE, HU6 C18 and HU14 C18, and PB C18 and PUB C18 (Figure 6A). When we ranked proteins based on the PC1 score, a sigmoidal-shaped plot was obtained (Figure 6B), with a long flat section extending across the majority of proteins (ca. protein rank 200–3500). This indicates that variability across methods was constant for the majority of proteins, while a few proteins had extremely low variability across methods (curve tail facing

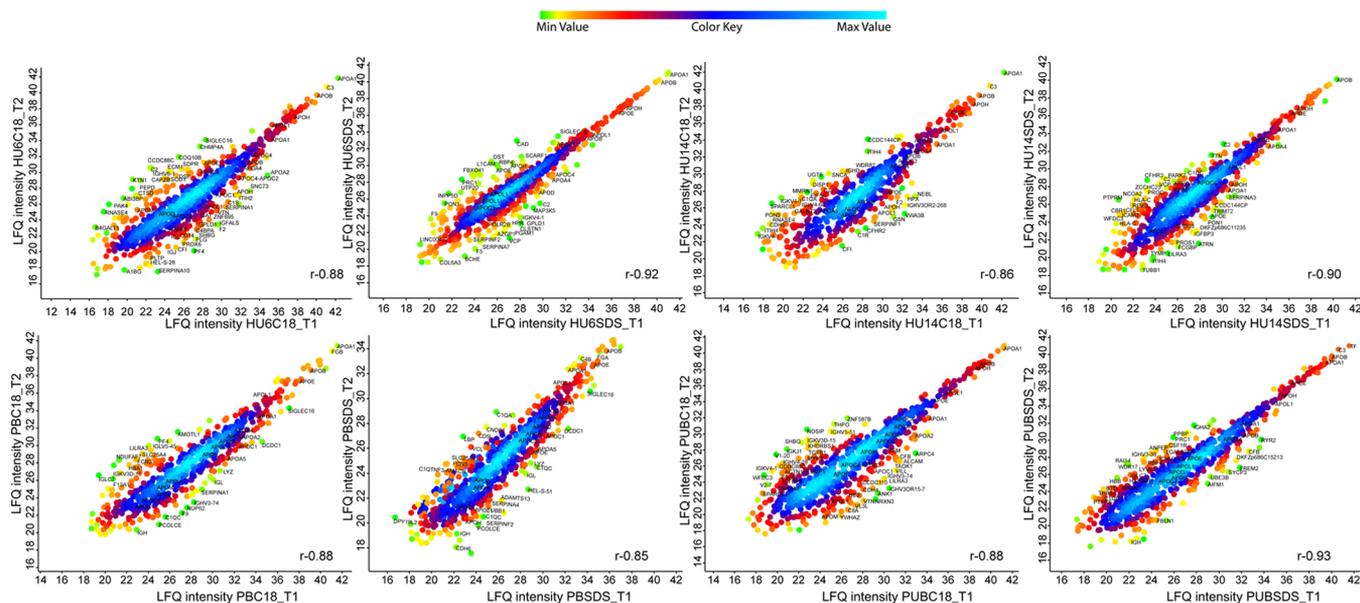


Figure 3. Density scatter plots of technical replicates for each of the eight methods, with a density color scale showing areas of higher and lower data density (blue and green, respectively). Pearson correlation results are indicated as r values on each plot.

upward on the far right), and a few proteins had higher variability across methods (curve tail facing downward on the far left). Members of two large families of proteins, apolipoproteins and the serpin proteases, are marked on the PC1 plot (Figure 6B), and we note that the level of PC1 variability across methods seems to be a function of protein concentration: the higher concentration plasma proteins (APOA2, APOC4-APOC2, APOC1, SERPINA1, SERPINF2, SERPINA6) are detected at the higher rank (right side of the curve) and the lower concentration plasma proteins (SERPINB1, SERPINE1, SERPINA10) appear in the lower rank (left side of the curve) in Figure 6B. We also ranked proteins for each depletion method based on their LFQ intensities together and separately (Figure 7) and annotated the plots with nine proteins whose protein concentration in plasma is known. This concentration range spans from low pg/mL to low mg/mL, representing ≥ 9 orders of magnitude of concentration. These proteins were identified in all eight individual workflows and were similarly distributed across the rank plots. Further, we compared the top 100 highest abundance plasma dynamic range of proteins in all depletion methods using Venn diagrams to find the number of overlapping proteins in C18 and SDS PAGE, and we found $\geq 70\%$ protein overlap except for HU14, which showed 57% (Figure S3).

Volcano Plot Analysis

To evaluate possible systematic bias across methods tested, we performed a volcano plot analysis for all combinations of methods, resulting in a total of 28 plots (Figure 8). We compared all of the methods to identify the variations in the statistically significant proteins using common proteins across methods. Using the 3064 proteins common across methods (Figure 2A), the UniProt-based KEGG pathways identified a total of 39 proteins, which were significantly differentially abundant across methods (Figure 8), representing $\sim 1\%$ of the full plasma proteome identified here. This small overall difference across methods is not too surprising given that an identical sample was analyzed across workflows and indicates

that significant quantitative differences attributable to the different sample workup methods are minimal. Furthermore, this number can be taken as a quasi-indicator of the protein quantification false discovery rate, and at $\sim 1\%$, it is encouragingly low. One of these proteins is complement C3, a protein depleted using the Hu14 column, and it is of interest that this is picked up as one of the 39 differentially expressed proteins since the difference in this case is likely real and attributable to the method used.

Human Protein Atlas-Based Tissue Annotation

To attempt an analysis of the likely tissue origins of the plasma proteins identified in our workflows, we searched for tissue specificity of our total 3064 commonly identified high-confidence proteins within the Human Protein Atlas (HPA) (<https://www.proteinatlas.org/>).^{45–47} We provide a list of 51 organs/tissues in which our 3064 HPA curated proteins can be found in Table S8. To obtain a global overview of the origin of our protein list, by specificity to organ systems, the 3064 proteins were collapsed into the 11 main organ systems of the human body (Figure 9A). Next, we sought to map brain-specific proteins using the human brain protein atlas. The analysis mapped an average of 1430 of our detected proteins, and their distribution in 10 subregions of the brain is shown in Figure 9B. There was an average of 14 region-specific proteins, while minimum protein numbers were found in the thalamus (1) and midbrain (2) and maximum were in the pons and medulla (39) and basal ganglia (28). Identification of tissue-specific proteins in our plasma proteome data set reflects its high potential for peripheral tissue and brain-specific biomarker identification in altered physiological conditions or in response to disease.

DISCUSSION

The eight methods of plasma high-abundance protein depletion and low-abundance protein enrichment evaluated in this study allow considerable enhancement of typical plasma proteome coverage, with a total number of 4385 proteins identified with high confidence using LC-MS/MS (data-

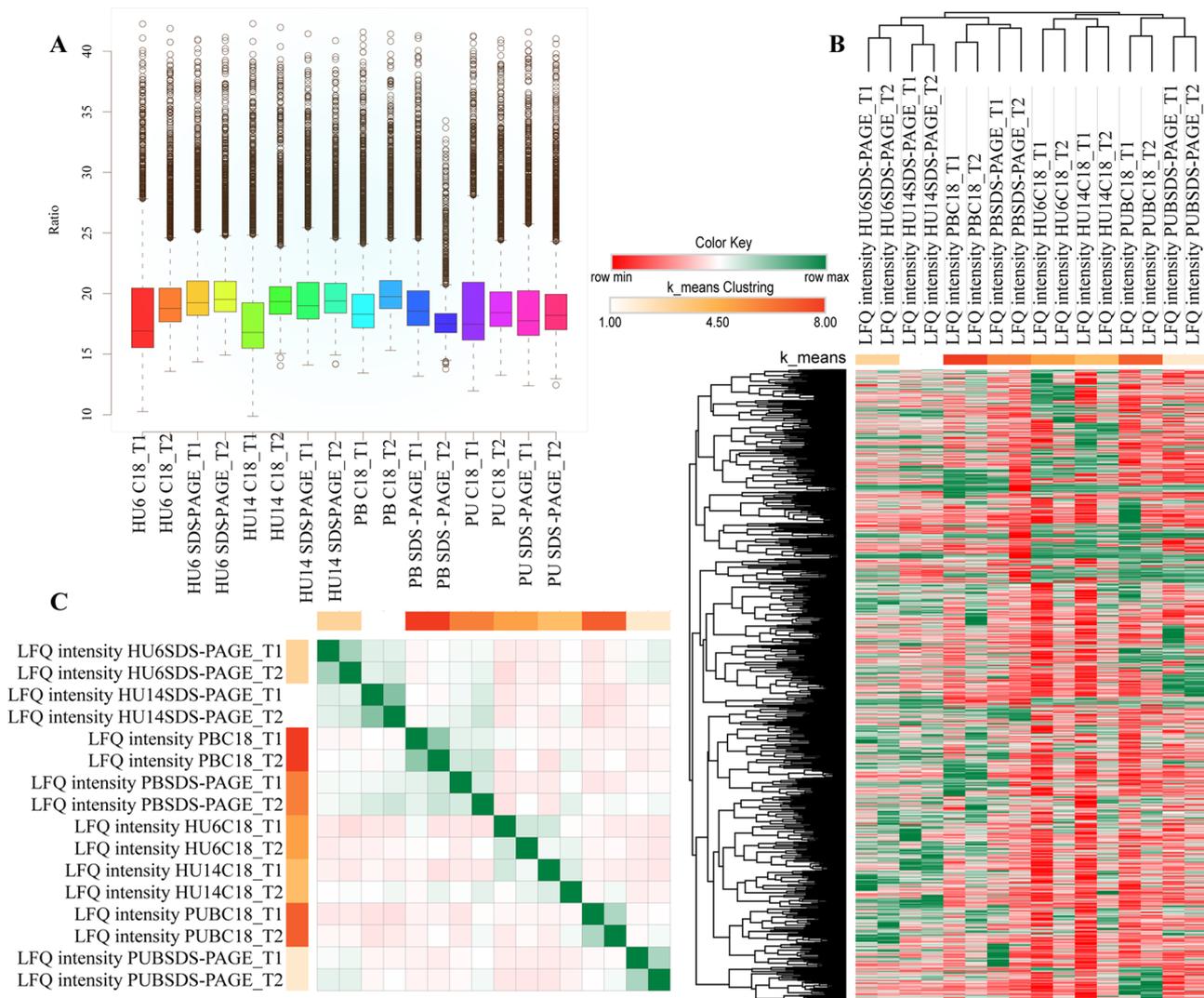


Figure 4. (A) Box-dot plots of LFQ intensity of all 16 experimental runs. The line within the box denotes the median value, and the upper and lower ranges of the box indicate the 5 and 95 percentiles of the intensities, respectively. The dots above the box plots are outliers. (B) Hierarchical cluster analysis and heat map for the 3064 proteins common to all methods. The *k* means algorithm was used to identify protein clusters. (C) The similarity matrix and heat map were constructed using the Pearson correlation values of the 16 methods, clustered based on the *k* means algorithm.

dependent analysis) and database searching with MaxQuant and Trans-Proteomic Pipeline. Plasma is arguably one of the most challenging sample types for this kind of analysis, and we have demonstrated here that good proteome coverage can be achieved even on this quantitatively and qualitatively complex proteome if appropriate sample processing approaches are applied. Our results compare well with other reported work, which has aimed to maximize plasma proteome coverage.^{16–28,48,49}

The human genome project has identified approximately 30 000 human genes (<https://www.genome.gov/human-genome-project>), while the human proteome draft article reported the identification of 17 294 proteins in total using mass spectrometry.⁵⁰ In our analysis, we identified a total of 4385 confidently identified proteins across all workflows, representing 3766, 3792, 3470, and 3682 proteins in the HU6, HU14, PB, and PUB workflows, respectively. Therefore, our estimate of the human proteome coverage, as a percentage of the 17 294 proteins reported by Kim et al.,⁵⁰ is as follows: total = 25.52%, HU6 = 21.78%, HU14 = 21.93%, PB = 20.06%, and

PUB = 21.29%. The average coverage we identified is 21.27% with standard error of the mean (SEM) = 0.42%.

Of the 4385 total proteins identified, 3064 (ca. 70%) overlap across methods, and excellent correlations are observed for all within-method technical replicates, and even reasonable correlations for across-method technical replicates were obtained. Consequently, we observe that all eight methods performed well in facilitating identification of high numbers of proteins, and all achieve comparable protein identifications to the highest of recent reports detailed in Table 2.^{16–28} A recent literature review observed that of ~180 plasma proteomics studies (conducted during 2005–2017), only 50% reported identification of 500 of the most abundant plasma proteins, following reanalysis of data sets using PeptideAtlas.⁵¹ This is likely a reflection of the enthusiasm for minimal sample processing since most approaches perform at most a single-step fractionation, usually focused on removal of one or more of the highest abundance proteins.

For initial fractionation, both high-abundance protein removal (Hu6, Hu14) and low-abundance protein enrichment (ProteMiner beads) perform well in reducing the plasma

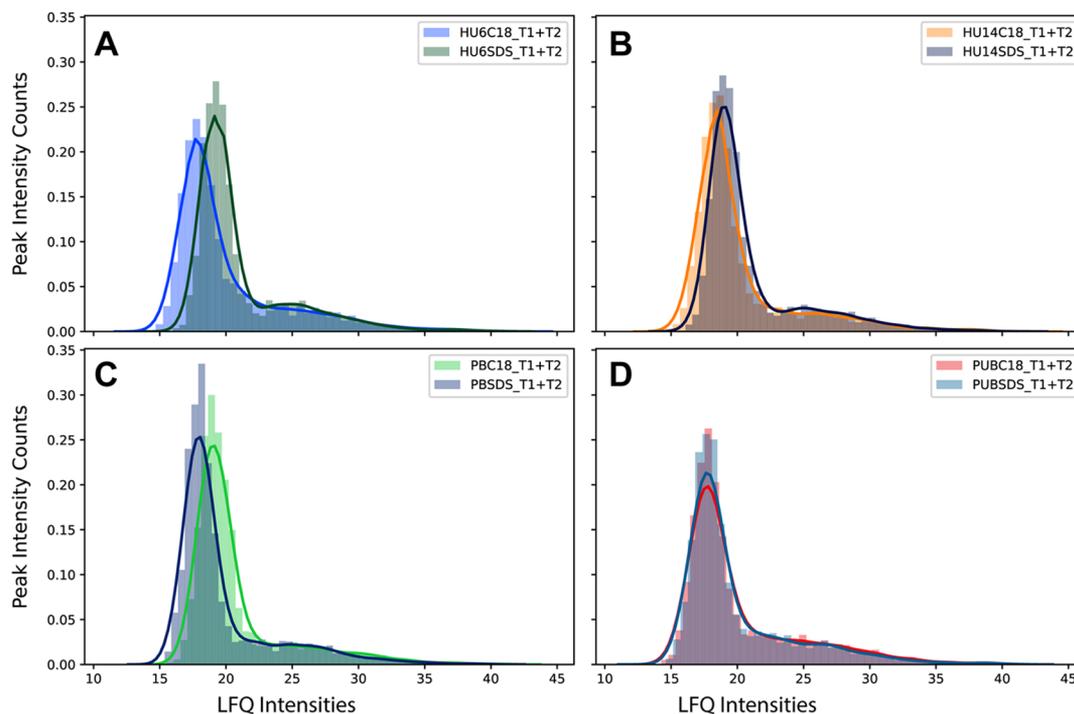


Figure 5. Juxtaposed LFQ intensities for pairwise comparison of C18 vs SDS PAGE fractionation using the following initial depletion approaches: (A) HU6, (B) HU14, (C) ProteoMiner bound (PB), and (D) ProteoMiner unbound (PUB).

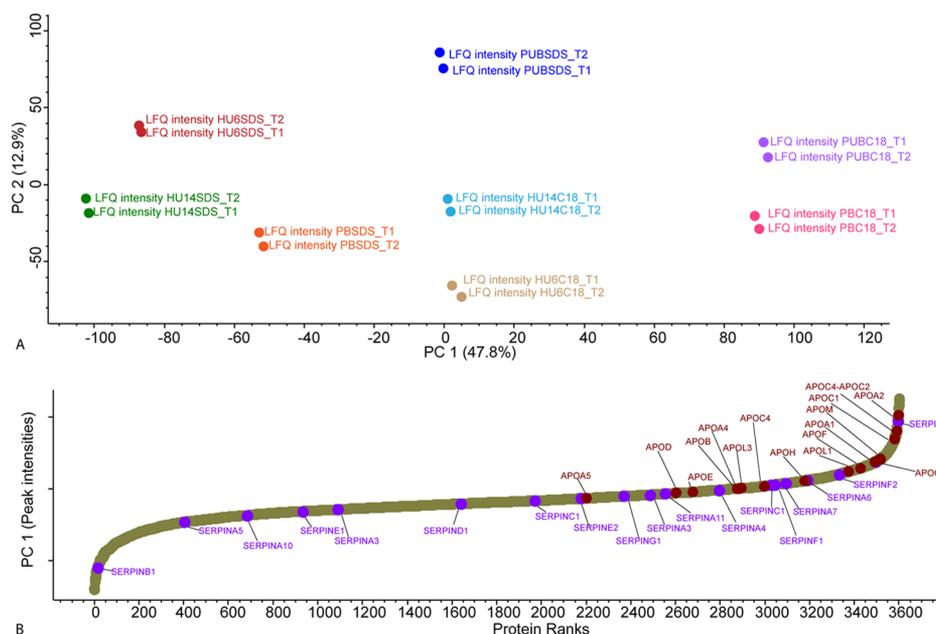


Figure 6. Principal component analysis: (A) two-dimensional (2D) principal component analysis using the first two principal components. (B) One-dimensional classification of all of the proteins identified in protein rank order based on the first principal component. The long flat section in the sigmoidal-shaped plot indicates that the variability across methods was constant for the majority of proteins. The level of PC1 variability across methods seems to be a function of protein concentration, with higher concentration plasma proteins, such as the majority of apolipoproteins, detected at the higher rank (right side of the curve) and lower concentration plasma proteins (SERPINB1, SERPINAS) appearing in the lower rank (left side of the curve). The apolipoproteins are annotated in red, and the SERPIN family proteins are annotated in purple.

protein dynamic concentration range. One advantage of high-abundance protein removal over low-abundance protein enrichment is that the antibody-based immunoaffinity chromatographic techniques (Hu6 and Hu14) are highly effective in the specific removal of the highest abundance proteins, with minimal nonspecific loss of other proteins.⁵² By comparison, the bead-based (ProteoMiner) method enriches

the lower abundance proteins but also loses much of the proteome to the “unbound” fraction (i.e., bead saturation), which risks loss of quantitative data (unless both the bound and unbound fractions are analyzed), thereby increasing the time and complexity for discovery-based quantitative projects. Previous studies have also shown the presence of almost equal numbers of proteins in bound and unbound fractions, and a

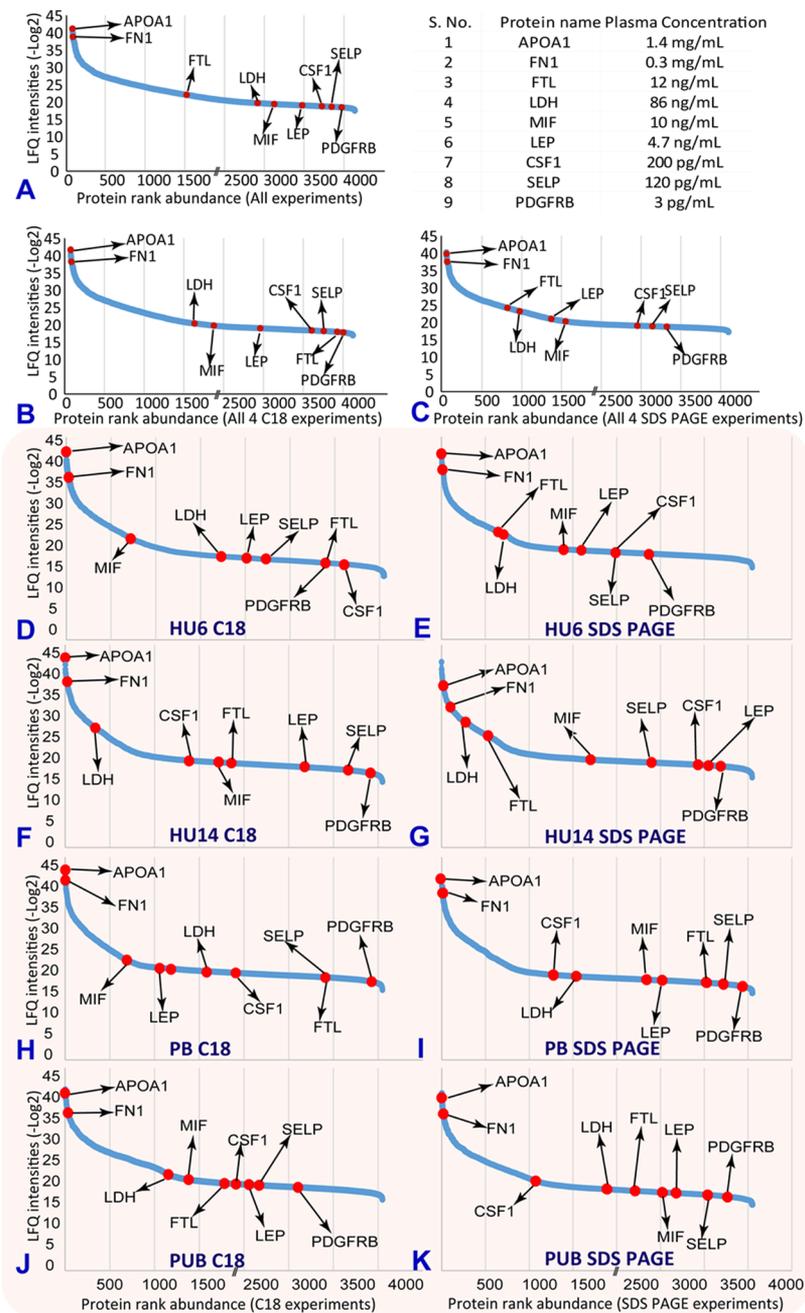


Figure 7. Protein rank abundance plot based on LFQ intensities. (A) All experiments together, (B) the four SDS PAGE methods, (C) the four C18 fractionation methods, and individual fractionation methods (D) HU6 C18, (E) HU6-SDS PAGE, (F) HU14 C18, (G) HU14-SDS PAGE, (H) PB C18, (I) PB-SDS PAGE, (J) PUB C18, and (K) PUB-SDS PAGE are presented in plots separately. The *x*-axis of the plot describes the protein rank, while the *y*-axis denotes Log₂ transformed LFQ intensities. Nine proteins whose protein concentration has previously been identified in the literature are marked on the plots, and their concentrations are reported in the table at the top right.

large overlap (91% identified proteins) was found between crude samples and depleted unbound fractions. Our data support these observations.⁵³ The fact that we identified such a high number of bound proteins is mainly due to the extensive fractionation strategies we applied post low-abundance protein enrichment. Most proteomics studies that have employed the ProteoMiner kit with no additional fractionation report identification of ~500–1000 proteins.^{16,19,20,54}

While analyzing the ProteoMiner unbound fractions (flowthrough), we identified 3682 total proteins in the flowthrough, indicating that the fractionation facilitated identification of increased numbers of proteins but also

demonstrated bead saturation of a very large number of proteins, which would be problematic for subsequent quantitative work. To overcome this quantification issue, it is possible that the proteins uniquely expressed in the bound fraction, but not present in the unbound fraction, could be compared in a relative quantification approach since these are not saturated. Then, the saturated proteins within the unbound fraction could be quantified separately, and in this case, the depletion via the bound fraction may even enhance quantitative differences in case vs control comparisons. However, this would add complexity to the methodology,

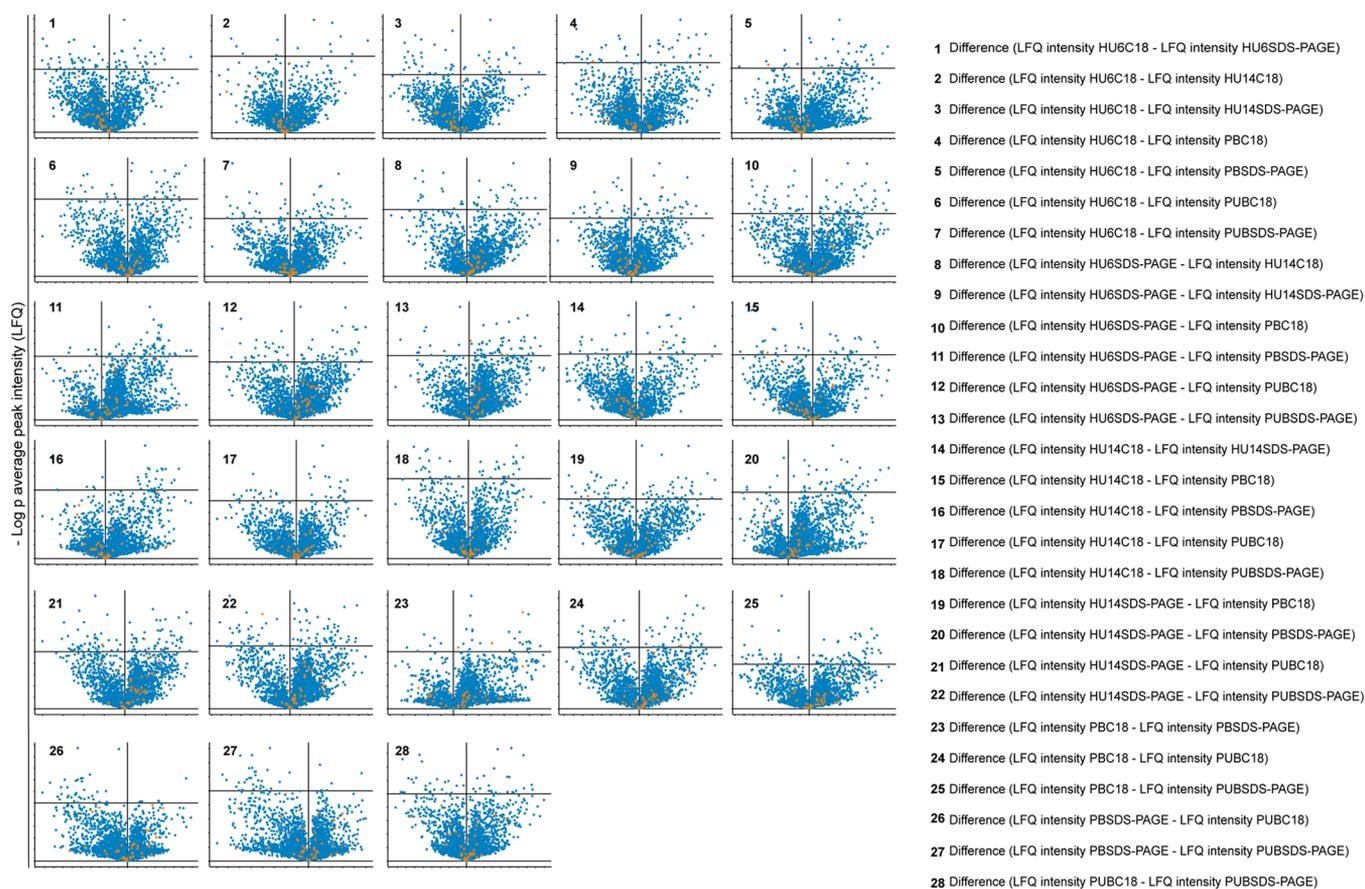


Figure 8. Volcano plot analysis: volcano plot showing significantly differentially abundant proteins in 28 comparisons of the tested methods. The plots show the difference of LFQ intensities between method pairs (*x*-axis) vs the $-\log_{10} p$ -values (*y*-axis) for each protein. The plots were based on $n = 3064$ proteins, and the horizontal bar in each plot represents $-\text{Log } p$ value 2.5 (i.e., ca. $p = 0.005$).

double the analysis time, and increase the cost since both fractions would need to be prepared and assayed.

In the second phase separations (C18 vs SDS PAGE), higher overall LFQ intensities, and in most cases also peak counts, were achieved in the Hu6 and Hu14 depleted samples, with the methods that included SDS PAGE fractionation. We are unclear why this should be; however, it is possible that the LC fractionation dilutes the peptide fractions, so peptide losses are likely to be greater as the diluted peptides bind nonspecifically to tube surfaces. By contrast, SDS PAGE does not dilute the sample and in fact traps and concentrates proteins within gel bands until they are liberated during in-gel tryptic digest. Throughout these steps, the sample volumes are minimal, providing fewer opportunities for nonspecific binding. It is likely that these problems with the C18 fractionation might be overcome using higher protein concentrations to start with and/or low protein binding tubes. Here, we used a starting amount of 50 μg of low-abundance plasma proteins for all methods to enable a direct comparison of the eight workflows. The amount of protein used reflects a level of low-abundance protein that can easily be recovered from a 50–100 μL aliquot of plasma. This is an important consideration for clinical and biobank samples, where plasma from population-based studies can be limited and in demand for a broad range of studies. Procedures that result in minimal sample loss are important in this context. Apart from minimizing the sample volume required while maximizing the data output, sample throughput is another important consideration if clinical

studies are to be completed in a reasonable timeframe. A further advantage of SDS PAGE is that it is inherently a higher throughput method than C18 LC because tens to thousands of samples can easily be run in parallel during the 1 h gel run time and in a relatively cost-effective manner since the equipment for SDS PAGE is considerably less costly than HPLC equipment. The C18 approach allows only serial separation, thereby greatly increasing the sample preparation time in direct proportion to the number of samples. Furthermore, equipment and running costs are considerably lower for SDS PAGE than for HPLC. Our observations are supported by other published work demonstrating a throughput advantage of gel-based separation vs LC fractionation.⁵⁵

The identification and quantification of over 10 000 proteins in samples like tissue and cell lines are possible without extensive prefractionation.⁵⁶ This is true because while cellular proteomes are complex in terms of numbers of proteins, their concentration range is much less so (likely 10^3 – 10^4).⁵⁷ By contrast, the concentration range of plasma proteins has been reported to be as high as 10^{13} ,⁴⁸ with just a handful of proteins representing >80% of the total protein content, while thousands of proteins are represented within the remaining <20% fraction. Consequently, to achieve protein coverage of the medium-/low-abundance proteins in plasma, particularly the lower abundance tissue-specific proteins that may reflect disease-specific change, plasma fractionation is unavoidable. Keshishian et al.²⁸ demonstrated this point by fractionating plasma and reported identification of ~ 5000 proteins from 16

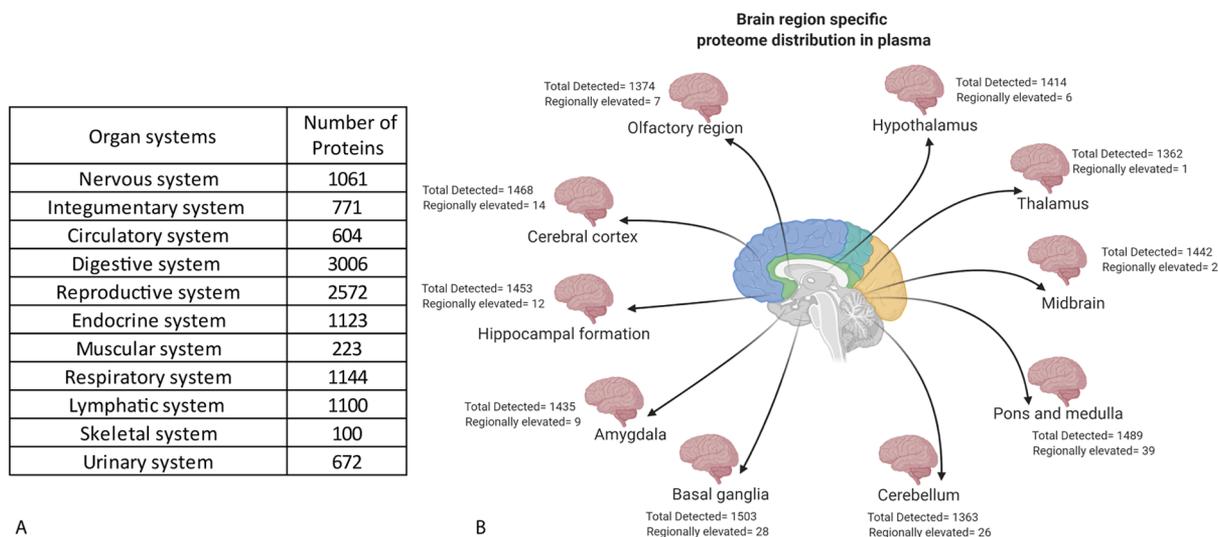


Figure 9. (A) Proteins identified in the plasma proteome, representative of the 11 main organ systems: nervous system (brain spinal cord, sensory organs), digestive system (mouth, oral mucosa, salivary gland, esophagus, intestine, stomach, colon, duodenum, gall bladder), muscular system (heart, skeletal, smooth), skeletal system (bone, cartilage, ligaments, tendons), integumentary system (skin, hair, nails, breasts, soft tissue, adipose tissue), reproductive system (ovaries, testis, vagina, uterus, cervix, endometrium, epididymis, fallopian tube, prostate, seminal vesicle), urinary system (kidneys, bladder), circulatory system (liver, blood, blood vessels, bone marrow), endocrine system (pituitary, thyroid, parathyroid, adrenal, pancreas), lymphatic system (lymph nodes, spleen, thymus, appendix, tonsil), and respiratory system (lungs, nasopharynx). Each of the proteins in our combined data set (either 3064 or 4385 proteins) was manually checked against the Human Protein Atlas (HPA) (<https://www.proteinatlas.org/>) to determine which proteins are the most abundantly expressed in 51 different organs (see Table S8 for a detailed breakdown of the protein numbers mapped to 51 specific organs). (B) Similarly, proteins from our plasma proteome were mapped to brain-specific subregions using Human Brain Atlas (<https://www.proteinatlas.org/humanproteome/brain>). The regionally enriched proteins that were identified in our plasma proteome are reported.

plasma samples using high-pH reversed-phase separation in combination with iTRAQ four-plex labeling. This work represents one of the most comprehensive single-study plasma proteomics profiling projects to date, utilizing a three-step procedure of fractionation, including (1) depletion of the top 14 HAPs (IgY14 LC20 column), (2) removal of the next ~50–100 moderate abundant proteins (MAPs) using a Sigma-Aldrich Supermix LC10 column, and (3) tryptic digestion of the remaining proteins and fractionating by C18 HPLC (28). This powerful approach was an excellent demonstration of the effectiveness of fractionation to yield high coverage of the plasma proteome. However, the methodology is also complex, expensive, time-consuming, and unlikely to be practical for many clinical projects, thereby restricting the general utility and replication of results in other laboratories. Consequently, cost-effective and easily accessible plasma proteomics analysis remains a challenge.

The work we present here demonstrates that simpler and faster approaches, accessible to most general biochemistry laboratories, can yield similarly high proteome coverage. Some examples of potential biomarkers that are usually present in plasma at ng/mL concentrations were identified in our data, including superoxide dismutase (SOD2), ribonuclease 4 (RAB4), 72 kDa type IV collagenase (matrix metalloproteinase 2 (MMP2)), and α -1-antitrypsin (SERPIN) proteins. Among these are some proteins with a role in disease, such as RAB4, which protects neurons from degeneration in amyotrophic lateral sclerosis (ALS) by stimulating neurofilament formation and protects hypothermia-induced degeneration in mouse embryonic cortical neurons.⁵⁸ Another protein identified in our data set is matrix metalloproteinase 2 (MMP2), which is involved in remodeling of vasculature, neural progenitor cell migration, and tumor invasion.⁵⁹ Some of the lowest low-

abundant plasma proteins that we identified include macrophage colony-stimulating factor 1 (200 pg/mL), P-selectin (120 pg/mL), platelet-derived growth factor receptor β (3 pg/mL), and leptin (4.7 ng/mL), some of which have also previously been reported.⁶⁰

It is evident from our study that a variety of different approaches can achieve depth of plasma proteome coverage (61). Two critical points appear to be (a) removal of the relatively few highest abundant proteins that represent >60–80% of the plasma total protein and (b) fractionation of the remaining medium-to-low-abundance protein fraction. Based on the eight methods we evaluated, we suggest the following: (1) with a view to subsequent quantification, high-abundance protein removal is preferable to low-abundance peptide enrichment to avoid the bead saturation problem; and (2) for high throughput where analysis of high numbers of clinical samples is required (tens to thousands), SDS PAGE is a faster approach than C18 as a secondary fractionation approach. Consequently, of the variety of methods evaluated here, the HU6-SDS PAGE and HU14-SDS PAGE approaches best satisfy all requirements, including narrowing the dynamic concentration range and the resultant depth of proteome coverage, retaining quantitative characteristics of the sample, and relatively simple, cost-effective, and time-efficient workflow. There is no doubt that plasma proteomics can provide high-confidence diagnostic tools for better stratification and monitoring of patient response to treatment.^{4,5,28,61}

CONCLUSIONS

In this study, we demonstrate that a variety of relatively simple prefractionation techniques can achieve depth of plasma proteome coverage. The results provide a clear demonstration of the reproducibility of these prefractionation techniques and

Table 2. Mass-Spectrometry-Based Plasma Proteomic Studies Allowing Comparison of Methodology, Run Time, and Numbers of Proteins Identified^a

title of the study	depletion strategy	mass spectrometry platform, data analysis software database/s used	number of fractions/samples × run time for each fraction/sample (min)	total run time (min)	total number of proteins	PRIDE accession
Affinity Capture Enrichment versus Affinity Depletion: A Comparison of Strategies for Increasing Coverage of Low-Abundant Human Plasma Proteins (ref 16)	affinity-based probes based on agarose-immobilized benzamide (ABA), <i>O</i> -phospho- <i>L</i> -tyrosine (pTYR), 8-amino-hexyl-cAMP (cAMP), or 8-amino-hexyl-ATP (ATP) HU14 (Agilent Technologies)	Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific, Bremen, Germany)	100 × 76	7600	total proteins = 1165	PXD020727
In-Depth Plasma Proteomics Reveals Increase in Circulating PD-1 during Anti-PD-1 Immunotherapy in Patients with Metastatic Cutaneous Melanoma (ref 17)	HU14 Agilent Technologies	Proteome Discoverer 2.4.0.305 (Thermo Fisher Scientific, Bremen, Germany) UniProt/SwissProt human database (January 2019)	72 × 50	3600	HU14 = 422; ProteomeMiner = 590; ABA = 598; ATP = 449; cAMP = 436; pTYR = 404 1917	PXD017201
Blood Plasma Proteomic Modulation Induced by Olanzapine and Risperidone in Schizophrenia Patients (ref 18)	HU14 Agilent Technologies	Q-Exactive mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) Galaxy-P project, including MSGF+ (v10072) and Percolator (v2.10), SpectrumAI pipeline human protein subset of ENSEMBL80	patients group A 14 × 39	patients group A = 546	patients group A = 319	PXD015049
Age-Dependent Changes in the Plasma Proteome of Healthy Adults (ref 19)	ProteoMiner (Bio-Rad Laboratories, Hercules, CA)	Synapt G2-Si mass spectrometer (Waters Corporation) Progenesis Q1 for Proteomics (QIP) software package with Apex3D, Peptide 3D, and Ion Accounting Informatics (Waters Corporation) UniProt human database (January 2019)	patients group B 40 × 39	patients group B = 1560	patients group B = 251	
Glyco-CPL: An Integrated Method for In-Depth and Comprehensive N-Glycoproteome Profiling of Human Plasma (ref 20)	ProteoMiner column (Bio-Rad Laboratories, Hercules, CA)	Orbitrap fusion lumos (Thermo Fisher) MaxQuant search engine UniProt human database	18 × 90	1620	1069	PXD016199
Identification of Potential Plasma Biomarkers for Abdominal Aortic Aneurysm Using Tandem Mass Tag Quantitative Proteomics (ref 21)	TOP 12 depletion kit (Pierce, Thermo Fisher)	Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher) MaxQuant search engine UniProt human database (version 2015_03)	6 × 78	468	healthy control = 567 patients = 537	PXD016428
Proteomes of Paired Human Cerebrospinal Fluid and Plasma: Relation to Blood–Brain	HU14 Agilent Technologies	Q-Exactive mass spectrometer (Thermo Fisher) Mascoat search engine version 2.3 (Matrix Science LTD, London, UK) UniProtKB/Trembl human database (December 2019) Ultimate 3000 RSLC nanosystem (Thermo Scientific)	27 × 90	2430	169	PXD011309
			120 × 150	18 000	422	PXD009589

Table 2. continued

title of the study	depletion strategy	mass spectrometry platform, data analysis software database/s used	number of fractions/samples × run time for each fraction/sample (min)	total run time (min)	total number of proteins	PRIDE accession
Barrier Permeability in Older Adults (ref 22)		Proteome Discoverer (version 1.4, Thermo Scientific)				
Plasma Proteome Profiles of Stable CAD Patients Stratified According to Total Apo C-III Levels (ref 23)	hybrid SPE(R)phospholipid, an RP-C4 cartridge for solid phase extraction (SPE)	UniProtKB/SwissProt human database (December 2014) 5600+ TripleTOF mass spectrometer system (AB Sciex, Concord, Canada)	52 × 17	884	188	PXD005973
Temporal Expression Profiling of Plasma Proteins Reveals Oxidative Stress in Early Stages of Type 1 Diabetes Progression (ref 24)	HU14 Agilent Technologies	Mascot v.2.4 (Matrix Science Inc., Boston, MA) and Protein Pilot software v4.2 (AB SCIEX, Concord, Canada) UniProt SwissProt human database (July 2015) Q-Exactive HF mass spectrometer (Thermo Fisher Scientific)	21 × 120	2520	2235	PXD007884
Discovery of a Potential Plasma Protein Biomarker Panel for Acute-on-Chronic Liver Failure Induced by Hepatitis B Virus (ref 25)	HU14 Agilent Technologies	MaxQuant software (v1.5.3.30) SwissProt human database (February 2016) nano-ESI Orbitrap Elite mass spectrometer (Thermo Fisher Scientific)	14 × 90	12.60	427	PXD007975
Proteomics Reveals the Effects of Sustained Weight Loss on the Human Plasma Proteome (ref 26)	HU6 Agilent Technologies followed by the Proteo-Prep20 plasma immunodepletion kit	MaxQuant search engine (v1.5.1.10) UniProt/SwissProt human database (IPLhuman.v3.87) Q-Exactive HF Orbitrap (Thermo Fisher Scientific)	52 × 45	23.40	1294	PXD004242
Plasma Proteome Profiling Reveals Dynamics of Inflammatory and Lipid Homeostasis Markers after Roux-En-Y Gastric Bypass Surgery (ref 27)	High Select Top14 Abundant Protein Depletion Midi Spin Column Multiple Affinity Removal Spin Cartridge Human 6	MaxQuant search engine (v1.5.3.23) Uniprot FASTA human database (June 2015) Q-Exactive HF Orbitrap (Thermo Fisher Scientific)	24 × 45	1080	1928	PXD009348
Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury (ref 28)	IgY14 LC20 followed by a Supermix LC10 column	MaxQuant search engine (v1.5.3.23) Uniprot FASTA human database (2011) Q-Exactive mass spectrometer (Thermo Fisher Scientific)	30 × 172	5160	4641	not submitted
		Spectrum Mill MS Proteomics Workbench software package v.4.2 β (Agilent Technologies)				27

Table 2. continued

title of the study	depletion strategy	mass spectrometry platform, data analysis software database/s used	number of fractions/samples × run time for each fraction/sample (min)	total run time (min)	total number of proteins	PRIDE accession
current study	1. Hu6 C18	Uniprot human database (2014) Q-Exactive plus mass spectrometer (Thermo Electron, Bremen, Germany), MaxQuant and TPP data analysis	24 × 60	1440	1. 3364	PXD02246
	2. Hu6-SDS PAGE	MaxQuant (version 1.6.10.43) and the Trans-Proteomic Pipeline (TPP) version 5.1.0 (November 2017)	(equal run time for all eight methods combination tested)		2. 3752	
	3. Hu14 C18	UniProt (SwissProt and TrEMBL) (2020)			3. 3778	
	4. Hu14-SDS PAGE				4. 3508	
	5. PB C18				5. 3421	
	6. PB-SDS PAGE				6. 3157	
	7. PUB C18				7. 3663	
	8. PUB-SDS PAGE				8. 3813	
					(proteins identified in all eight methods combination)	

^aThe studies were compared using mass spectrometer run time for determining the depth of proteome coverage.

show that gel-based approaches can be an alternative to expensive and time-consuming chromatographic separation, thereby significantly accelerating the analysis time. It is encouraging that a variety of methodologies can achieve similarly high proteome coverage, which allows flexibility of choice depending on project-specific needs, such as whether qualitative or quantitative information is sought, size of the project in terms of sample numbers, availability of specific lab resources, etc. These considerations are important in the effort to accelerate plasma proteomics research so as to provide efficient, reliable, and accurate diagnoses for the population as a whole.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00670>.

Representative NuPAGE LDS gel profile of undepleted plasma and depleted plasma containing low-abundant plasma proteins and high-abundant plasma proteins depleted of high-abundance proteins using HU6, HU14, and ProteoMiner depletion methods (Figure S1); multiscatter plot depicting the cross-correlation of all of the methods (Figure S2); Venn diagram showing an overlap of the top 100 dynamic range of proteins between depletion methods HU6 C18 and HU6-SDS PAGE, HU14 C18 and HU14-SDS PAGE, PB C18 and PB-SDS PAGE, and PUB C18 and PUB-SDS PAGE (Figure S3) (PDF)

Full table describing the identification of the analyzed MaxQuant results of HU6-SDS PAGE and C18 (Table S1), HU14-SDS PAGE and C18 (Table S2), PB-SDS PAGE and C18 (Table S3), and PUB-SDS PAGE and C18 (Table S4); full list of protein identifications using the Trans Proteomics Pipeline (TPP), resulting from use of three search engines (Mascot, Comet, and XTandem), and combining all eight sample types (HU6-SDS PAGE and C18, HU14-SDS PAGE and C18, PB-SDS PAGE and C18, PUB-SDS PAGE and C18), consisting of 24 fractions and 2 technical replicates and resulting in 384 raw files (Table S5); MaxQuant parameter set and parameter set for TPP (Table S6); MaxQuant-derived information on the raw peptide matches for each of the four high-abundance protein removal steps (Table S7); detailed breakdown of the protein numbers mapped to 51 specific organs separately from the Human Protein Atlas (HPA) (<https://www.proteinatlas.org/>) (Table S8) (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Anne Poljak – Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, NSW 2052, Australia; Bioanalytical Mass Spectrometry Facility, Mark Wainwright Analytical Centre, University of New South Wales, Sydney, NSW 2052, Australia; Phone: +61 (2) 9385 3950; Email: a.poljak@unsw.edu.au

Authors

Gurjeet Kaur – Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, NSW 2052, Australia; Bioanalytical Mass Spectrometry Facility,

Mark Wainwright Analytical Centre, University of New South Wales, Sydney, NSW 2052, Australia; orcid.org/0000-0002-5125-0531

Syed Azmal Ali – Cell Biology and Proteomics Lab, National Dairy Research Institute, Karnal, Haryana 132001, India

Ling Zhong – Bioanalytical Mass Spectrometry Facility, Mark Wainwright Analytical Centre, University of New South Wales, Sydney, NSW 2052, Australia

Mark J. Raftery – Bioanalytical Mass Spectrometry Facility, Mark Wainwright Analytical Centre, University of New South Wales, Sydney, NSW 2052, Australia

Perminder Sachdev – Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, NSW 2052, Australia; Neuropsychiatric Institute, Euroa Centre, Prince of Wales Hospital, Sydney, NSW 2052, Australia

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00670>

Author Contributions

A.P. and P.S. conceived the project; G.K. and A.P. designed the experiments; G.K. performed the chromatographic and SDS/PAGE fractionation experiments and preparation of peptides for LC-MS/MS; A.P. acquired the LC-MS/MS data with assistance from L.Z. and M.J.R. regarding instrument setup, parameter settings, and advice on optimizing sample preparation steps; G.K. and S.A.A. conducted bioinformatics (MaxQuant and TPP) data analysis; G.K., A.P., and S.A.A. wrote and revised the manuscript; and P.S. supervised the project and reviewed the manuscript. All authors read, edited, and revised the manuscript and accepted the final version.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

G.K. is thankful to the UNSW Scientia fellowship. We would also like to thank the Bioanalytical Mass Spectrometry Facility, University of New South Wales, for providing the necessary facilities to carry out the research work.

■ REFERENCES

- (1) Végvári, A.; Welinder, C.; Lindberg, H.; Fehniger, T. E.; Markovarga, G. Biobank resources for future patient care: developments, principles and concepts. *J. Clin. Bioinf.* **2011**, *1*, No. 24.
- (2) Anderson, N. L.; Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **2002**, *1*, 845–867.
- (3) Rifai, N.; Gillette, M. A.; Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **2006**, *24*, 971–983.
- (4) Geyer, P. E.; Holdt, L. M.; Teupser, D.; Mann, M. Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* **2017**, *13*, No. 942.
- (5) Anderson, L. Six decades searching for meaning in the proteome. *J. Proteomics* **2014**, *107*, 24–30.
- (6) Anderson, N. L.; Anderson, N. G. The Human Plasma Proteome. *Mol. Cell. Proteomics* **2002**, *1*, 845.
- (7) Omenn, G. S. Exploring the human plasma proteome. *Proteomics* **2005**, *5*, 3223–3225.
- (8) Cox, J.; Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **2011**, *80*, 273–299.
- (9) Geiger, T.; Cox, J.; Mann, M. Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol. Cell. Proteomics* **2010**, *9*, 2252–2261.

- (10) Muñoz, J.; Heck, A. J. From the human genome to the human proteome. *Angew. Chem., Int. Ed.* **2014**, *53*, 10864–10866.
- (11) Sturgeon, C.; Hill, R.; Hortin, G. L.; Thompson, D. Taking a new biomarker into routine use—a perspective from the routine clinical biochemistry laboratory. *Proteomics: Clin. Appl.* **2010**, *4*, 892–903.
- (12) Vihko, P.; Kontturi, M.; Korhonen, L. K. Purification of human prostatic acid phosphatase by affinity chromatography and isoelectric focusing. Part I. *Clin. Chem.* **1978**, *24*, 466.
- (13) Pieper, R.; Su, Q.; Gatlin, C. L.; Huang, S. T.; Anderson, N. L.; Steiner, S. Multicomponent immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics* **2003**, *3*, 422–432.
- (14) Qian, W. J.; Kaleta, D. T.; Petritis, B. O.; Jiang, H.; Liu, T.; Zhang, X.; et al. Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy. *Mol. Cell. Proteomics* **2008**, *7*, 1963–1973.
- (15) Tu, C.; Rudnick, P. A.; Martinez, M. Y.; Cheek, K. L.; Stein, S. E.; Slebos, R. J.; et al. Depletion of abundant plasma proteins and limitations of plasma proteomics. *J. Proteome Res.* **2010**, *9*, 4982–4991.
- (16) Palström, N. B.; Rasmussen, L. M.; Beck, H. C. Affinity Capture Enrichment versus Affinity Depletion: A Comparison of Strategies for Increasing Coverage of Low-Abundant Human Plasma Proteins. *Int. J. Mol. Sci.* **2020**, *21*, No. 5903.
- (17) Babačić, H.; Lehtiö, J.; de Coaña, Y. P.; Pernemalm, M.; Eriksson, H. In-depth plasma proteomics reveals increase in circulating PD-1 during anti-PD-1 immunotherapy in patients with metastatic cutaneous melanoma. *J. Immunother. Cancer* **2020**, *8*, No. e000204.
- (18) Garcia-Rosa, S.; Carvalho, B. S.; Guest, P. C.; Steiner, J.; Martins-de-Souza, D. Blood plasma proteomic modulation induced by olanzapine and risperidone in schizophrenia patients. *J. Proteomics* **2020**, No. 103813.
- (19) Xu, R.; Gong, C. X.; Duan, C. M.; Huang, J. C.; Yang, G. Q.; Yuan, J. J.; Zhang, Q.; Xiong, X. Y.; Yang, Q. W. Age-Dependent Changes in the Plasma Proteome of Healthy Adults. *J. Nutr., Health Aging* **2020**, 846.
- (20) Zhang, Y.; Mao, Y.; Zhao, W.; Su, T.; Zhong, Y.; Fu, L.; Zhu, J.; Cheng, J.; Yang, H. Glyco-CPLL: an integrated method for in-depth and comprehensive N-glycoproteome profiling of human plasma. *J. Proteome Res.* **2020**, *19*, 655–666.
- (21) Henriksson, A. E.; Lindqvist, M.; Sihlbom, C.; Bergström, J.; Bylund, D. Identification of Potential Plasma Biomarkers for Abdominal Aortic Aneurysm Using Tandem Mass Tag Quantitative Proteomics. *Proteomes* **2018**, *6*, No. 43.
- (22) Dayon, L.; Cominetti, O.; Wojcik, J.; Galindo, A. N.; Oikonomidi, A.; Henry, H.; Migliavacca, E.; Kussmann, M.; Bowman, G. L.; Popp, J. Proteomes of paired human cerebrospinal fluid and plasma: relation to blood–brain barrier permeability in older adults. *J. Proteome Res.* **2019**, *18*, 1162–1174.
- (23) Manfredi, M.; Chiariello, C.; Conte, E.; Castagna, A.; Robotti, E.; Gosetti, F.; Patrone, M.; Martinelli, N.; Bassi, A.; Cecconi, D.; Marengo, E.; Olivieri, O. Plasma Proteome Profiles of Stable CAD Patients Stratified According to Total Apo C-III Levels. *Proteomics: Clin. Appl.* **2019**, *13*, No. 1800023.
- (24) Liu, C. W.; Bramer, L.; Webb-Robertson, B. J.; Waugh, K.; Rewers, M. J.; Zhang, Q. Temporal expression profiling of plasma proteins reveals oxidative stress in early stages of Type 1 Diabetes progression. *J. Proteomics* **2018**, *172*, 100–110.
- (25) Zhou, N.; Wang, K.; Fang, S.; Zhao, X.; Huang, T.; Chen, H.; Yan, F.; Tang, Y.; Zhou, H.; Zhu, J. Discovery of a potential plasma protein biomarker panel for acute-on-chronic liver failure induced by hepatitis B virus. *Front. Physiol.* **2017**, *8*, No. 1009.
- (26) Geyer, P. E.; Wewer Albrechtsen, N. J.; Tyanova, S.; Grassl, N.; Iepsen, E. W.; Lundgren, J.; Madsbad, S.; Holst, J. J.; Torekov, S. S.; Mann, M. Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Mol. Syst. Biol.* **2016**, *12*, No. 901.
- (27) Albrechtsen, N. J. W.; Geyer, P. E.; Doll, S.; Treit, P. V.; Bojsen-Møller, K. N.; Martinussen, C.; Jørgensen, N. B.; Torekov, S. S.; Meier, F.; Niu, L.; Santos, A.; et al. Plasma proteome profiling reveals dynamics of inflammatory and lipid homeostasis markers after Roux-en-Y gastric bypass surgery. *Cell Syst.* **2018**, *7*, 601.e3–612.e3.
- (28) Keshishian, H.; Burgess, M. W.; Gillette, M. A.; Mertins, P.; Clauser, K. R.; Mani, D. R.; Kuhn, E. W.; Farrell, L. A.; Gerszten, R. E.; Carr, S. A. Multiplexed, quantitative workflow for sensitive biomarker discovery in plasma yields novel candidates for early myocardial injury. *Mol. Cell. Proteomics* **2015**, *14*, 2375–2393.
- (29) Simpson, R. J. Staining proteins in gels with coomassie blue. *Cold Spring Harbor Protoc.* **2007**, *2007*, No. pdb.prot4719.
- (30) Shevchenko, A.; Tomas, H.; Havli, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006**, *1*, 2856–2860.
- (31) Gundry, R. L.; White, M. Y.; Murray, C. I.; Kane, L. A.; Fu, Q.; Stanley, B. A.; Van, E. Y. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Curr. Protoc. Mol. Biol.* **2010**, *90*, 10.25.1–10.25.23.
- (32) Prikas, E.; Poljak, A.; Ittner, A. Mapping p38 α mitogen-activated protein kinase signaling by proximity-dependent labeling. *Protein Sci.* **2020**, *29*, 1196–1210.
- (33) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **2016**, *11*, 2301.
- (34) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **2011**, *10*, 1794–1805.
- (35) Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **2014**, *13*, 2513–2526.
- (36) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372.
- (37) Tyanova, S.; Temu, T.; Sinitcyn, P.; et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **2016**, *13*, 731–740.
- (38) Suhail, A.; Rizvi, Z. A.; Mujagond, P.; Ali, S. A.; Gaur, P.; Singh, M.; Ahuja, V.; Awasthi, A.; Srikanth, C. V. DeSUMOylase SENP7-Mediated Epithelial Signaling Triggers Intestinal Inflammation via Expansion of Gamma-Delta T Cells. *Cell Rep.* **2019**, *29*, 3522–3538.
- (39) Chopra, A.; Ali, S. A.; Bathla, S.; Rawat, P.; Vohra, V.; Kumar, S.; Mohanty, A. K. High-Resolution Mass Spectrometer–Based Ultra-Deep Profile of Milk Whey Proteome in Indian Zebu (Sahiwal) Cattle. *Front. Nutr.* **2020**, No. 150.
- (40) Bjornson, R. D.; Carriero, N. J.; Colangelo, C.; et al. X! Tandem, an improved method for running X!tandem in parallel on collections of commodity computers. *J. Proteome Res.* **2008**, *7*, 293–299.
- (41) Stanley, J. R.; Adkins, J. N.; Slysz, G. W.; et al. A statistical method for assessing peptide identification confidence in accurate mass and time tag proteomics. *Anal. Chem.* **2011**, *83*, 6135–6140.
- (42) Shteynberg, D.; Deutsch, E. W.; Lam, H.; et al. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol. Cell. Proteomics* **2011**, *10*, No. M111.007690.
- (43) Chen, Y. Y.; Dasari, S.; Ma, Z. Q.; Vega-Montoto, L. J.; Li, M.; Tabb, D. L. Refining comparative proteomics by spectral counting to account for shared peptides and multiple search engines. *Anal. Bioanal. Chem.* **2012**, *404*, 1115–1125.
- (44) Kapp, E. A.; Schutz, F.; Connolly, L. M.; Chakel, J. A.; Meza, J. E.; Miller, C. A.; Fenyo, D.; Eng, J. K.; Adkins, J. N.; Omenn, G. S.; Simpson, R. J. An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. *Proteomics* **2005**, *5*, 3475–3490.
- (45) Uhlén, M.; Fagerberg, L.; Hallström, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.;

Asplund, A.; Olsson, I.; et al. Tissue-based map of the human proteome. *Science* **2015**, *347*, No. 1260419.

(46) Saha, S.; Harrison, S. H.; Shen, C.; Tang, H.; Radivojac, P.; Arnold, R. J.; Zhang, X.; Chen, J. Y. HIP 2: An online database of human plasma proteins from healthy individuals. *BMC Med. Genomics* **2008**, *1*, No. 12.

(47) Nanjappa, V.; et al. Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Res.* **2014**, *42*, D959–D965.

(48) Schwenk, J. M.; Omenn, G. S.; Sun, Z.; Campbell, D. S.; Baker, M. S.; Overall, C. M.; Aebersold, R.; Moritz, R. L.; Deutsch, E. W. The human plasma proteome draft of 2017: building on the human plasma PeptideAtlas from mass spectrometry and complementary assays. *J. Proteome Res.* **2017**, *16*, 4299–4310.

(49) Li, N.; Zhou, Y.; Wang, J.; Niu, L.; Zhang, Q.; Sun, L.; Ding, X.; Guo, X.; Xie, Z.; Zhu, N.; Zhang, M.; et al. Sequential Precipitation and Delipidation Enables Efficient Enrichment of Low-Molecular Weight Proteins and Peptides from Human Plasma. *J. Proteome Res.* **2020**, *19*, 3340–3351.

(50) Kim, M. S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; Thomas, J. K. A draft map of the human proteome. *Nature* **2014**, *509*, 575–581.

(51) Farrah, T.; Deutsch, E. W.; Omenn, G. S.; et al. A High-Confidence Human Plasma Proteome Reference Set with Estimated Concentrations in PeptideAtlas. *Mol. Cell. Proteomics* **2011**, *10*, No. M110.006353.

(52) Muenchhoff, J.; Poljak, A.; Thalamuthu, A.; Gupta, V. B.; Chatterjee, P.; Raftery, M.; Masters, C. L.; Morris, J. C.; Bateman, R. J.; Fagan, A. M.; Martins, R. N.; Sachdev, P. S. Changes in the plasma proteome at asymptomatic and symptomatic stages of autosomal dominant Alzheimer's disease. *Sci. Rep.* **2016**, *6*, No. 29078.

(53) Jankovska, E.; Svitek, M.; Holada, K.; Petrak, J. Affinity depletion versus relative protein enrichment: a side-by-side comparison of two major strategies for increasing human cerebrospinal fluid proteome coverage. *Clin. Proteomics* **2019**, *16*, No. 9.

(54) Chen, L.; Zhao, W.; He, J.; Li, L.; Meng, D.; Cai, D.; Yu, J.; Chen, D.; Wu, Y.; Zhou, T. Label-Free Quantitative Proteomic Screening of Candidate Plasma Biomarkers for the Prognosis of Breast Cancer with Different Lymph Node Statuses. *Proteomics: Clin. Appl.* **2018**, *12*, No. e1700117.

(55) Choksawangkar, W.; Edwards, N.; Wang, Y.; Gutierrez, P.; Fenselau, C. Comparative study of workflows optimized for in-gel, in-solution, and on-filter proteolysis in the analysis of plasma membrane proteins. *J. Proteome Res.* **2012**, *11*, 3030–3034.

(56) Geiger, T.; Wehner, A.; Schaab, C.; Cox, J.; Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **2012**, *11*, No. M111.014050.

(57) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198–207.

(58) Li, S.; Sheng, J.; Hu, J. K.; et al. Ribonuclease 4 protects neuron degeneration by promoting angiogenesis, neurogenesis, and neuronal survival under stress. *Angiogenesis* **2013**, *16*, 387–404.

(59) Wang, L.; Zhang, Z. G.; Zhang, R. L.; et al. Matrix Metalloproteinase 2 (MMP2) and MMP9 Secreted by Erythropoietin-Activated Endothelial Cells Promote Neural Progenitor Cell Migration. *J. Neurosci.* **2006**, *26*, 5996.

(60) Ignjatovic, V.; Geyer, P. E.; Palaniappan, K. K.; et al. Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. *J. Proteome Res.* **2019**, *18*, 4085–4097.

(61) Surinova, S.; Schiess, R.; Hüttenhain, R.; Cerciello, F.; Wollscheid, B.; Aebersold, R. On the development of plasma protein biomarkers. *J. Proteome Res.* **2011**, *10*, 5–16.