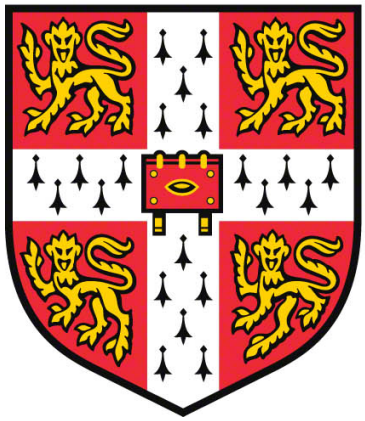
Development and validation of blood-based proteomic biomarker-sociodemographic diagnostic prediction models to identify major depressive disorder among symptomatic individuals



Sung Yeon Han

Trinity College

Department of Chemical Engineering and Biotechnology

University of Cambridge

This thesis is submitted for the degree of *Doctor of Philosophy*.

September 2020

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

Sung Yeon Han

September 2020

Development and validation of blood-based proteomic biomarker-sociodemographic diagnostic prediction models to identify major depressive disorder among symptomatic individuals

Sung Yeon Han

Abstract

Major depressive disorder (MDD) is a highly prevalent and disabling condition with a complex pathophysiology that has not been fully elucidated to date. While the socioeconomic burden of the disease is significant, many individuals remain undiagnosed or misdiagnosed. This is largely because the current diagnostic approach that relies on clinical evaluations of signs and symptoms can be subjective, and time and resources tend to be rather limited in primary care where the majority seek help for depression. Therefore, there is a significant and pressing need for an objective, reliable and readily accessible diagnostic test to enable earlier and more accurate diagnosis of MDD. In particular, as individuals experiencing subthreshold levels of depressive symptoms have an increased risk of developing MDD, it would be clinically relevant for such a diagnostic test to be able to identify depressed patients and/or individuals with high risks of incident MDD among symptomatic individuals.

This thesis sought to develop risk prediction models that could potentially be utilised within a clinical setting to facilitate earlier and more accurate diagnosis of MDD. Such models were used to obtain probability estimates of the investigated individuals having or developing MDD based on their blood-based proteomic profiles and other characteristics, including sociodemographic and lifestyle factors. A targeted mass spectrometry approach was used to measure the abundances of a panel of peptides representing proteins, many of which have been previously associated with psychiatric disorders. Biomarkers were investigated in serum samples, which are widely used for blood-based biomarker discovery, as well as in dried blood spot samples, which are relatively novel in the field and carry several advantages. Importantly, this thesis focused on adopting appropriate statistical methods to ensure that the diagnostic predictions made by the models were accurate and reproducible, by addressing problems of

model overfitting and model selection uncertainty. A particularly significant aspect of this was the development and application of a multimodel-based approach combining feature extraction and model averaging, which resulted in improved model predictive performance and generalisability.

Diagnostic prediction models based on serum proteomic, sociodemographic/lifestyle and clinical data were shown to be able to differentiate between subthreshold symptomatic individuals who developed and did not develop MDD. Additionally, diagnostic prediction models based on dried blood spot proteomic and digital mental health assessment data were shown to be able to identify currently depressed patients without an existing MDD diagnosis as well as currently not depressed patients with an existing MDD diagnosis among subthreshold symptomatic individuals. These results clearly demonstrate the potential of such prediction models to be used as an aid to the diagnosis of MDD in clinical practice, especially within the primary care setting. Moreover, MDD was found to be associated with several blood-based proteomic biomarkers, which mainly represented an immune/inflammatory profile, as well as with various other patient features, most notably body mass index and childhood trauma. Although further investigations are needed, these associations reveal disturbances in the stress response pathways involving the hypothalamic-pituitary-adrenal axis in the pathophysiology of depression.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor, Prof. Sabine Bahn, without whom this thesis would not be possible. I wish to thank her for giving me the opportunity to undertake this PhD at the Cambridge Centre for Neuropsychiatric Research and for helping me to grow as a research scientist. Her guidance, support and advice throughout the past four years have been invaluable, and I am also grateful to her for generously funding a part of my PhD. My gratitude extends to Trinity College and the Department of Chemical Engineering and Biotechnology for the support that I have received during my time as a student in Cambridge.

I would like to thank everyone who has contributed to the work presented in this thesis. This includes my statistical collaborators at the Cambridge Centre for Neuropsychiatric Research, whose supervision and guidance on the statistical aspects of my PhD were instrumental in the planning and execution of the projects that were carried out in this thesis. Further, the work of my experimental collaborators on mass spectrometry provided me with valuable datasets for the analyses conducted in this thesis. I am grateful to our external collaborators for making available the clinical samples and datasets that were investigated, and to all participants who took part in the studies and enabled this research to be possible. I would also like to acknowledge the Stanley Medical Research Institute for funding the various research projects that were carried out as part of my PhD.

Finally, my deepest gratitude goes to my fiancé, my parents and my future in-laws. I would not have been able to complete this journey without their unconditional love, support and care during these past years, and I thank them for encouraging me to pursue my academic interests in science. I am immensely grateful in particular to my fiancé, Kwan Woo, who has always been by my side and given me constant moral support throughout my studies, and especially in the writing of this thesis.

Publications

R. L. Preece*, **S. Y. S. Han***, and S. Bahn, "Proteomic approaches to identify blood-based biomarkers for depression and bipolar disorders," *Expert Rev. Proteomics*, vol. 15, no. 4, pp. 325–340, 2018.

*authors contributed equally

J. D. Cooper, **S. Y. S. Han**, J. Tomasik, S. Ozcan, N. Rustogi, N. J. M. van Beveren, F.M. Leweke, S. Bahn, "Multimodel inference for biomarker development: an application to schizophrenia," *Transl. Psychiatry*, vol. 9, no. 1, p. 83, 2019.

S. Y. S. Han, J. D. Cooper, S. Ozcan, N. Rustogi, B. W. J. H. Penninx, and S. Bahn, "Integrating proteomic, sociodemographic and clinical data to predict future depression diagnosis in subthreshold symptomatic individuals," *Transl. Psychiatry*, vol. 9, no. 1, p. 277, 2019.

T. Olmert, J. D. Cooper, **S.Y.S. Han**, G. Barton-Owen, L. P. Farrag, E. Bell, L. V. Friend, S. Thiahulan, S. Ozcan, N. Rustogi, R. L. Preece, P. Eljasz, J. Tomasik, D. Cowell, S. Bahn, "A Combined Digital and Biomarker Diagnostic Aid for Mood Disorders: The Delta Trial, an Observational Study," *JMIR Res. Protoc.*, vol. 9, no. 6, e18453, Aug. 2020.

S. Y. S. Han, J. Tomasik, N. Rustogi, S. G. Lago, G. Barton-Owen, P. Eljasz, J. D. Cooper, S. Ozcan, T. Olmert, L. P. Farrag, L. V. Friend, E. Bell, G. Thomas, R. Tuytten, S. Bahn., "Diagnostic prediction model development using data from dried blood spot proteomics and a digital mental health assessment to identify major depressive disorder among individuals presenting with low mood," *Brain. Behav. Immun.*, vol. 90, pp. 184–195, 2020.

Contents

List of Figures

List of Tables

Abbreviations

A1AG1	α-1-acid glycoprotein 1
A2GL	Leucine-rich α-2-glycoprotein
AACT	α-1-antichymotrypsin
AIC	Akaike information criterion
AL1A1	Retinal dehydrogenase 1
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
APOE	Apolipoprotein E
APOH	Apolipoprotein H
AUC	Area under the ROC curve
BD	Bipolar disorder
BDNF	Brain-derived neurotrophic factor
BMI	Body mass index
CFAH	Complement factor H
CIDI	Composite Interview Diagnostic Instrument
CNS	Central nervous system
CRF	Corticotropin-releasing factor
CRP	C-reactive protein
CSF	Cerebrospinal fluid
CV	Coefficient of variation
DALY	Disability-adjusted life year
DBS	Dried blood spot
DNA	Deoxyribonucleic acid
DSM	Diagnostic and Statistical Manual of Mental Disorders
ELISA	Enzyme-linked immunosorbent assay
EPV	Events per variable
FDA	Food and Drugs Administration
FETUA	Fetuin-A
FPR	False positive rate
GP	General practitioner

GWAS	Genome-wide association study
HBA	Haemoglobin subunit α
HPA	Hypothalamus-pituitary-adrenal
ICD	International Statistical Classification of Diseases and Related Health Problems
IDS	Inventory of Depressive Symptomatology
IL	Interleukin
IRRID	International Registered Report Identifier
LASSO	Least absolute shrinkage and selection operator
LC	Liquid chromatography
m/z	Mass-to-charge ratio
MAOI	Monoamine oxidase inhibitor
MDD	Major depressive disorder
MRM	Multiple reaction monitoring
MS	Mass spectrometry
NEO FFI	NEO Five-Factor Inventory
NESDA	Netherlands Study of Depression and Anxiety
PC	Principal component
PCA	Principal component analysis
PHLD	Glycoprotein phospholipase D
PHQ	Patient Health Questionnaire
QC	Quality control
QIDS	Quick Inventory of Depressive Symptomatology
QQQ	Triple quadrupole
ROC	Receiver operating characteristic
SIL	Stable isotope-labelled
TCA	Trycyclic antidepressant
TNF	Tumour necrosis factor
TPR	True positive rate
WEMWBS	Warwick-Edinburgh Mental Well-Being Scale
WHO	World Health Organization

Remaining abbreviations of protein names can be found in the **Appendix** (**Table A.1**).

Chapter 1 Introduction

1.1. Major depressive disorder

1.1.1. Burden and challenges

Major depressive disorder (MDD; also referred to as depression) is a complex, heterogeneous and burdensome disorder, which is characterised by recurrent (one or more) episodes of low mood and energy levels over a prolonged period. Other symptoms include loss of interest or pleasure; changes in weight or appetite; sleeping problems; feelings of worthlessness or guilt; concentration or decision-making difficulties; recurrent thoughts of death or suicide; psychomotor agitation or retardation [1]. MDD is the most common psychiatric disorder, with more than 300 million people affected worldwide (around 4% of the global population) [2]. It has a lifetime prevalence of 16% [3], affecting approximately one in five women and one in eight men during their lifetime [4], and a 12-month prevalence of 6% [5].

MDD is a highly disabling condition that can cause the affected individuals to suffer from considerably reduced quality of life and functional impairment at work, school and in the family [6]. It also has various negative consequences for physical health, as the risks of heart disease, hypertension, stroke, diabetes, Alzheimer's disease, cancer and obesity are increased in those with MDD compared to those without MDD [7]. Depression increases the risk of all-cause mortality by 60-80%, and is a major contributor to deaths by suicide [8]–[10]. A meta-review estimated that the risk of suicide is almost 20-fold higher in MDD patients compared with the general population [10]. The World Health Organization (WHO) predicts that by 2030, MDD will become the most debilitating disorder worldwide (overtaking cardiovascular disease) [11], and the largest contributor to the global burden of disease, accounting for 7.5% of all disability-adjusted life-years (DALYs; years of healthy life lost due to disability and premature mortality) [2]. In addition, the direct and indirect economic costs relating to MDD are substantial. The total annual costs of MDD in Europe have been estimated at €92 billion (26% direct healthcare costs, 15% direct non-medical costs, 59% indirect costs) [12].

Despite the significant psychological, social and economic burden of depression, many people

remain undiagnosed or misdiagnosed, and therefore untreated or inadequately treated [13]. This is not least as the aetiological and pathophysiological mechanisms which underpin the disease remain to be fully elucidated, but also as the translation of research findings into clinical practice is challenging [14], [15]. Effective clinical mangement of depression is hindered by several factors, including: the subjective nature of diagnosis; 'trial-and-error' approach to drug treatment selection; heterogeneous clinical presentation (symptomatology), disease course and treatment response; high comorbidity and overlapping symptoms with other mental disorders; inconsistent research findings; lack of a clear molecular understanding of disease; lack of adequate animal models; lack of resources and health-care providers; and social stigma associated with mental disorders.

1.1.2. Diagnosis

Currently, the diagnosis of MDD is based on the clinical evaluation of self-reported symptoms by a psychiatrist or other healthcare professional. The diagnostic criteria, including the number and duration of symptoms required, are outlined in formal classification systems, such as the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) [1] and the International Statistical Classification of Diseases and Related Health Problems, 11th Revision (ICD-11) [16]. For instance, according to the DSM-5, a diagnosis of MDD requires an individual to present with at least five out of nine depressive symptoms, including at least one core symptom of depressed mood or anhedonia (loss of interest or pleasure); the symptoms need to be present most of the day, nearly every day, for at least two weeks, and represent a change from previous functioning [1]. These criteria are summarised in **Table 1.1**. To date, no objective tests exist for the diagnosis of MDD (or any other psychiatric disorder) due to the limited biological understanding of the disease.

The majority of care for depression is delivered by general practitioners (GPs) within the primary care setting. However, clinical evaluations by GPs are usually time-constrained (between five and seven minutes on average, which are likely to be insufficient to reveal mental health problems [14]), and can also be subjective as their compliance with the formal diagnostic framework can vary depending on their respective clinical experience. Consequently, under-, over-, or misdiagnosis of depression are common problems in primary care. Based on a meta-analysis of over 50,000 patients, only about 47% of MDD patients are correctly identified by GPs, while 19% of non-depressed individuals are incorrectly identified [17]. Under-diagnosis

can be detrimental to the course and outcome of the illness, as patients are likely to receive no or inappropriate treatment, often for several years [18]. Fewer than 50% of those affected in the world and fewer than 10% in some countries, especially low-income countries, receive effective treatment such as pharmacotherapy or psychotherapy [11], [19]. Moreover, under-diagnosis could also arise from the difficulties associated with recognising depression in patients who present mainly with somatic symptoms, such as lack of energy and general aches and pains, which is the case for approximately two thirds of patients with depression in primary care [20], [21]. As GPs may become more focused on investigating a potential underlying organic condition in such patients, their mental health problems could consequently be overlooked. Additionally, the patients themselves may also resist a diagnosis of depression due to the belief that their symptoms could be related to physical causes, which is particularly problematic in certain cultures that associate mental illnesses with stigma, discrimination and social exclusion [14]. On the other hand, over-diagnosis (misidentification) of MDD can lead to over-treatment [17]. In particular, inappropriate treatment with antidepressants may be burdensome, associated with an unnecessary risk of side effects, as well as potentially stigmatising [22], [23]. A study of over 5,000 participants with clinician-identified depression in the community found that while the majority of the participants were being prescribed and using psychiatric medications, only around 40% met the DSM-IV criteria for experiencing a major depressive episode (MDE) in the past 12 months when assessed using a structured interview [24].

In order to improve the diagnostic reliability, clinical research (and post-mortem) studies often use the WHO World Mental Health Composite Interview Diagnostic Instrument (CIDI) to determine the diagnoses of psychiatric disorders [25]. The CIDI is a fully structured, comprehensive and standardised diagnostic interview which is designed to be used by trained lay interviewers to assess mental disorders in accordance with the DSM-IV and ICD-10 criteria. Diagnoses generated by the CIDI generally show good concordance with those based on clinician-administered structured diagnostic interviews [26], [27]. Although CIDI was originally intended to be used in psychiatric epidemiological studies, it is also widely used in the clinic as well as in research [28].

Table 1.1. DSM-5 criteria for the diagnosis of MDD.

Note that criteria A-C represent a MDE. Abbreviations: DSM (Diagnostic and Statistical Manual of Mental Disorders); MDE (major depressive episode); MDD (major depressive disorder).

A. Five (or more) of the following symptoms have been present during the same two-week period and represent a change from previous functioning, and at least one of the symptoms is either (1) or (2). 1) Depressed mood most of the day, nearly every day 2) Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day 3) Significant weight loss or gain, or decrease or increase in appetite nearly every day 4) Insomnia or hypersomnia nearly every day 5) Psychomotor agitation or retardation nearly every day 6) Fatigue or loss of energy nearly every day 7) Feelings of worthlessness or excessive or inappropriate guilt nearly every day 8) Diminished ability to think or concentrate, or indecisiveness, nearly every day 9) Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide
B. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
C. The episode is not attributable to the physiological effects of a substance or to another medical condition.
D. The occurrence of the MDE is not better explained by schizoaffective disorder, schizophrenia, schizophreniform disorder, delusional disorder, or other specified and unspecified schizophrenia spectrum and other psychotic disorders.
E. There has never been a manic episode or a hypomanic episode.

1.1.3. Disease course

The average age of MDD onset is around 25 years in both men and women, and the risk of MDD generally decreases with age after early adulthood [5], [29]. Women have a two-fold increased risk of developing MDD after puberty than men, which reflects the gender difference in the disease prevalence [30].

The course of MDD is variable, such that in the general population, 50% of MDD patients

recover within three months (*i.e.*, the average duration of a MDE is three months), 63% within six months, and 76% within 12 months [31]. Psychiatric comorbidities are common in MDD, with 59% of patients with lifetime MDD having anxiety disorders [32], 44% having personality disorders [33], and 14% having substance use disorders [34]. The presence of psychiatric comorbidities leads to greater disease severity, less favourable disease course, and increased costs of MDD care [35]. Duration of current episode, symptom severity, and childhood trauma are also associated with lower recovery rates of MDD [1], [35], [36].

Moreover, MDD is a highly recurrent disorder. There is a 50% chance that individuals experience at least one more episode following their first episode, and the chance of further relapse increases to 70% and 90% following the second and third episodes, respectively [37]. The persistence of depressive symptoms during remission is a strong predictor of recurrence [1]. In addition, early onset depression (before the age of 20) significantly increases the risk of recurrence [38].

1.1.4. Treatment

Antidepressant medications are currently the first-line treatment for MDD. There are many types of antidepressants, including selective serotonin reuptake inhibitors, serotonin and noradrenaline reuptake inhibitors, tricyclic antidepressants (TCAs), monoamine oxidase inhibitors (MAOIs), and atypical antidepressants. Although there are over 25 antidepressants that have been approved by the Food and Drugs Administration (FDA) for the treatment of MDD, the overall effectiveness of antidepressant treatment has been suboptimal [39], [40]. A large meta-analysis of 182 clinical trials reported that the average response rate for antidepressants is only 54% (37% for placebo response) [41]. Such variable response to antidepressants is due to not only individual differences in drug metabolism and drug therapeutic efficacy, but also the disease heterogeneity of MDD and the existence of patient subgroups [42]. In the absence of biological evidence to aid in selecting among existing medications, the process of finding an effective treatment for a given patient is largely by 'trial-and-error', whereby various drugs of the same or different class are tested until symptom remission is achieved [42]. The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, which investigated more than 400 MDD patients, showed that only one-third of MDD patients achieved remission after an initial 12-week antidepressant treatment trial, and one-third required up to four consecutive antidepressant treatment trials to achieve remission,

whereas the remaining one-third did not achieve remission even after four trials [43]. This trial-and-error process of drug treatment selection can take up to two years, or even longer, until the patient recovers [44]. A prolonged treatment period increases the duration of depressive episodes and healthcare costs, and can lead to poorer outcomes [40]. The importance of early intervention in MDD is highlighted by the evidence of an inverse relationship between the duration of episode and treatment outcome [45], [46]. In addition, antidepressant treatment can lead to a range of unpleasant side effects, including nausea, increased appetite, weight gain, insomnia and fatigue, and increase the risk of suicidal thoughts and acts, particularly in adolescents and young adults [47].

Psychological and psychosocial interventions are also effective in treating MDD and can be used as an alternative to or in combination with pharmacotherapy. The most common types of psychotherapy for MDD are cognitive-behavioural therapy, behavioural activation therapy, psychodynamic therapy, interpersonal therapy, and problem-solving therapy [48]. There is growing evidence that combined treatment with psychotherapy and antidepressant medication can be more effective in relieving symptoms of depression than treatment with medication alone, and also enable better adherence to medication [49], [50]. It is usually recommended that patients experiencing mild depressive episodes are initially treated with psychotherapy, whereas those experiencing moderate and severe depressive episodes are treated with medication or a combination of medication and psychotherapy [51], [52]. The choice of intervention strategy, that is, pharmacotherapy, psychotherapy, or a combination of both, is made by GPs or clinicians following the consideration of several factors, including symptom profile, disease severity, patient preference, prior treatment history, and family treatment history [53]. Individuals with depression usually desire outcomes beyond symptom remission, which include a return to normal functioning and positive mental health [54].

1.1.5. Subthreshold depression

In recent years, the notion of subthreshold depression has received increased attention [47], [55]–[62]. This aims to recognise those individuals experiencing depressive symptoms that do not fulfil the full diagnostic criteria for MDD with respect to the number, severity and/or duration of symptoms (*i.e.*, fewer than five symptoms and/or duration of symptoms for less than two weeks and/or lack of a core symptom based on the DSM-5), and are consequently overlooked by the current checklist diagnostic approach. The UK National Institute for Health

and Care Excellence (NICE) has recently updated their guideline to define 'subthreshold depressive symptoms' as 'at least one key symptom of depression but with insufficient other symptoms and/or functional impairment to meet the criteria for full diagnosis' [52].

However, as an officially agreed definition of subthreshold depression has not yet been established, this condition has been labelled by a range of terminology across different studies, including subsyndromal depression, subclinical depression and minor depression. While the operational definitions used across the studies also vary in relation to the number and duration of symptoms required, most studies have defined subthreshold depression as experiencing two or more depressive symptoms, including at least one of depressed mood or anhedonia, which are the core symptoms of depression according to the DSM-5 criteria, for a minimum duration of two weeks [58]. Moreover, the reported prevalence rates of subthreshold depression range between 2.9% and 9.9% in primary care setting, and between 1.4% and 17.2% in community (general population) setting [63]–[66]. Nevertheless, studies have collectively shown that the presence of subthreshold depressive symptoms is associated with increased functional and social impairment, reduced quality of life, and increased utilisation of health services and economic costs [58], [60], [64], [67], [68]. The clinical relevance of subthreshold depression is also demonstrated by the similarities in demographic and clinical characteristics that are observed between those with depressive disorders and those with subthreshold depression, including the gender difference in the prevalence, the percentage of family history of depression, and the pattern of medical and psychiatric comorbidity [69]. A systematic review of the definitions, prevalence rates and factors associated with subthreshold depression can be found in *Rodríguez et al. (2012)* [58].

Longitudinal studies have investigated the progression of subthreshold depression into a full-blown depressive disorder, and the reported rates of individuals with subthreshold depression developing MDD in the general population range between 3.8% and 18.9% [55], [67], [70]–[73]. Across the studies, stricter definitions of subthreshold depression and longer follow-up periods were generally associated with higher rates of MDD reported at follow-up [62]. Importantly, subthreshold depression has been identified as a risk factor for incident MDD, such that those with subthreshold depression are more likely to develop MDD in the future compared to those without depressive symptoms [55], [72], [74], [75]. According to a systematic review of prospective studies, the relative risk of MDD among individuals with subthreshold depression is higher in general medical populations and in high-risk groups than

in the general population [67]. Additionally, a recent meta-analysis of 16 longitudinal studies revealed that the risk of developing MDD is approximately two times greater in individuals with subthreshold depression relative to non-depressed individuals, with subgroup analyses showing similar estimates of relative risk across different age groups (youth, adults and the elderly) as well as sample types (general community and primary care) [76]. Various risk factors have been identified for the development of MDD among those with subthreshold depression, including emotional neglect during childhood, lower perceived social support, suicidal ideation, sleeping difficulties, having recurrent short episodes of depression, the presence of an anxiety disorder, a substance use disorder or a chronic physical disorder, and lower mental or physical functioning [62], [73], [77]. The identification of recurrent short episodes of depression lasting at least three days, rather than at least one episode lasting at least two weeks, as a significant risk factor for MDD may be especially relevant considering that many previous studies on subthreshold depression have required a minimum symptom duration of two weeks, which means that those with recurrent short episodes of depression who are particularly at risk of developing MDD would have been disregarded [62].

Furthermore, there is growing evidence that proactively treating subthreshold individuals with indicated preventive interventions can help to prevent or delay the onset of MDD [78]–[80]. A meta-analysis of 32 randomised controlled trials found that preventive psychological interventions led to a 21% reduction in the incidence of MDD compared with controls, clearly demonstrating the feasibility of prevention of MDD onset [78]. Effective prevention strategies include reducing depressive symptoms before they fulfil the full diagnostic criteria for MDD, as well as increasing social support [48]. While it is estimated that current interventions only reduce about one-third of the disease burden associated with MDD [81], preventing or delaying its onset may contribute to the further reduction of the disease burden. However, as many individuals with subthreshold depression naturally recover and no longer report depressive symptoms at follow-up [62], [72], it is vital to ensure that those presenting with normal, self-remitting depressive symptoms are not given unnecessary interventions, which could not only be burdensome and potentially stigmatising [22], [23], but also result in a misallocation of scarce resources [82]. Hence, these findings highlight the clinical importance of identifying subthreshold symptomatic individuals who have particularly high risks of progressing to full-blown MDD, and thereby would benefit most from indicated preventive measures in primary care settings.

In addition, the notion of subthreshold depression is consistent with the view that depressive disorders exist along a dimensional contiuum of symptomatic severity rather than as distinct disease entities [83]–[86]. According to this view, which is also referred to as the spectrum view of depression, major depression and no depressive symptoms would lie at either ends of the spectrum and subthreshold depression would lie in between the two. Whereas the importance of subthreshold conditions is typically underestimated in a categorical approach to the classification of mental disorders as their impact on the lives of the affected individuals are overlooked, the presence of an arbitrary threshold is eliminated in such a dimensional approach, which consequently allows for a better identification of milder conditions of the disorders [58]. A study that investigated different categories of depressive disorders, namely MDD with seven to nine symptoms, MDD with five or six symptoms, subthreshold depression (two to four symptoms, including at least one key symptom), one key symptom, and no symptoms, found that more severe depression was associated with greater functional disability and healthcare utilisation [77]. In addition, the risk of developing MDD was found to be higher in those with subthreshold depression than in those with one key symptom, and lowest in those with no depressive symptoms [77]. Therefore, these findings support the spectrum view of depression, although it would still be important to consider qualitative differences between major depression and subthreshold depression [87].

1.1.6. Disease heterogeneity

The aetiology and clinical presentation of MDD are complex and heterogeneous. Based on the current symptoms-based approach to diagnosis, there are 1,497 unique combinations of symptoms that fulfil the diagnostic criteria for MDD under the DSM-5 [88]. Additionally, two individuals both diagnosed with MDD could have no symptoms in common [89], and such heterogeneity explains why the same treatment strategy yields variable outcomes in patients. Moreover, symptoms of MDD are often present in other mental disorders, mainly persistent depressive disorder (previously known as dysthymia), bipolar disorder (BD) and schizophrenia, which can hinder correct diagnosis [1], [48]. Hence, the establishment of strict boundaries between distinct diagnostic categories of mental disorders is challenging, as they often share overlapping symptoms as well as common genetic predispositions [90]. This is further complicated by the fact that the categorisation based on clinical observations of signs and symptoms does not align well with findings from clinical neuroscience and genetics, suggesting that it may not accurately reflect the underlying aetiological and pathophysiological

mechanisms of mental disorders [91]–[93]. As such, the Research Domain Criteria project, launched by the US National Institute of Mental Health, aims to improve the clinical management of mental disorders by incorporating physiological data into future classification schemes [93]. Relatedly, a long-standing debate in the field of psychiatry is that, rather than classifying mental disorders as distinct disease entities in a traditional categorical approach, it may be more appropriate to conceptualise them in a dimensional approach [91], [94].

1.1.7. Digital healthcare

In light of the considerable burden of mental disorders, the unmet need for effective clinical management, and the lack of adequate healthcare services [13], there is an increased interest in fostering technological advances to develop novel digital tools that can potentially improve psychiatric care [95]. Digital tools, in the form of web- or smartphone-based applications, enable increased capacity for data collection, user accessibility, as well as cost-efficiency [96]. For instance, digital adaptations of existing self-reported questionnaires are increasingly employed for the assessment of mental disorders, including depression [97]. While the transformation of established instruments into the digital format could potentially affect the reliability and validity of the results, as responses could vary depending on factors such as the presentation layout and the perceived sense of security and anonymity [98], a systematic review has found that the results obtained by the digital versions are generally comparable to those obtained by the traditional pen-and-paper versions [99].

Digital psychiatric assessments provide an increased convenience for both patients and clinicians/healthcare providers, as they can be completed by patients in the comfort of their homes without needing to visit the clinic or health centre, which may be particularly beneficial to those who are reluctant to seek help regarding their mental health. This also means that when patients do visit the clinic or health centre, consultation times can be saved and thereby spent more efficiently [100]. In addition to the digitalised versions of established paper instruments, new instruments are being developed specifically to be utilised online. These can have additional technological advantages, such as the personalisation of questions to individual users (*i.e.*, skipping irrelevant questions based on previous answers), as well as the inclusion of audio and video [97], [101]–[103]. Similarly, web- or smartphone-based interventions could offer a more time-efficient, accessible and cost-effective alternative to face-to-face interventions [103]. Therefore, digital tools represent a promising area of innovation and have the potential to

improve the clinical management of MDD and other mental disorders by facilitating earlier and/or more accurate diagnosis as well as supporting more effective interventions.

1.2. Pathophysiology of MDD

Despite advances in our understanding of the neurobioloy of MDD, there is currently no single theory that can account for all aspects of the disease. In fact, several hypotheses have emerged to explain the mechanisms that underlie the development of MDD [104], [105]. MDD appears to be a multifactorial disorder that arises from complex interactions between genetic predispositions, molecular and functional disturbances, and environmental factors [48], [106].

1.2.1. Genetics

The contribution of genetic factors to the risk of MDD is supported by converging evidence from family and twin studies [105]. The risk of MDD is three-fold higher in first-degree relatives of individuals with MDD compared with that in the general population, and the heritability rate of MDD is estimated to be approximately 37% [107], [108]. In addition, heritability appears to be significantly higher in women than in men [109], which supports the higher incidence and prevalence of depression in women.

Despite the considerable heritability of MDD, genome-wide association studies (GWAS) have had limited success in revealing consistent significant genetic effects and providing insights into the biological pathways involved in MDD [110]. For example, no significant findings were reported in a mega-analysis of 18,759 individuals (9,240 cases and 9,519 controls) by the Psychiatric Genomics Consortium [111], or in a meta-analysis of depressive symptoms of 34,549 individuals by the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium [112]. This is not only because the risk of MDD is highly polygenic and involves many susceptibility genes with small effect sizes (in other words, each susceptibility gene makes only a small contribution to the total genetic risk), but also given the complex genetic heterogeneity of the disorder, very large sample sizes are required to detect significant associations [48], [113]. More recently, GWAS involving larger sample sizes have been successful in identifying significant genetic variants. A meta-analysis of 461,134 individuals (135,458 cases and 344,901 controls) by the Psychiatric Genomics Consortium found 44

independent loci associated with clinically evaluated and self-reported MDD [113], and a meta-analysis of 807,553 individuals (246,363 cases and 561,190 controls) found 102 independent variants associated with depression, of which 87 were replicated in an independent sample of 1,306,354 individuals (414,055 cases and 892,299 controls) [114].

1.2.2. Environmental factors

Several environmental and sociodemographic factors are associated with an increased risk of MDD. These include childhood trauma (sexual, physical or emotional abuse during childhood) [115]–[117], the absence of a partner (including being divorced or widowed) [5], recent negative life events (such as financial problems and illness or loss of close relatives or friends) [118], unemployment, low educational attainment and low socioeconomic status [48], [119]. Lifestyle factors including smoking [120], [121], alcohol use [122]–[124], drug use [125]–[127], low physical activity [128], [129], and obesity [130]–[132] also increase the risk of MDD.

1.2.3. Monoamines

The monoamine hypothesis of depression proposes that depression is caused by a functional deficiency of the monoamine neurotransmitters serotonin, noradrenaline and/or dopamine in the central nervous system (CNS) [133], [134]. This was first implicated after the antihypertensive drug reserpine, which reduces the level of these monoamine neurotransmitters, was found to induce depression, and further supported by clinical observations that antidepressant medications which increase monoamine transmission, including TCAs and MAOIs, alleviate symptoms of depression [135]–[137]. As such, the field of pharmacotherapy of MDD continues to be dominated by monoamine-based compounds to this day [48]. Despite this, studies that measured noradrenaline and serotonin metabolites in the plasma, urine, cerebrospinal fluid (CSF) and brains (post-mortem) of MDD patients have revealed inconsistent findings, and some argue that the monoamine hypothesis of MDD may be overly simplistic [48], [138].

1.2.4. HPA axis

The hypothalamus-pituitary-adrenal (HPA) axis is a major neuroendocrine system that involves the hypothalamus, pituitary and adrenal glands, and plays an important role in stress response. Corticotropin-releasing factor (CRF; also known as corticotropin-releasing hormone)

is released from the hypothalamus in response to stress, which, in turn, stimulates the release of glucocorticoids such as cortisol from the adrenal glands in a negative feedback mechanism (feedback inhibition). Abnormalities of the HPA axis, including an impaired glucocorticoid-mediated feedback regulation, have been widely implicated in the pathophysiology of depression [48], [139]. Studies have found that MDD is associated with elevated levels of cortisol and CRF, and reduced glucocorticoid receptor sensitivity (also known as glucocorticoid resistance), as demonstrated by non-suppression of cortisol release in the dexamethasone test [140]–[143]. Hippocampal volume reductions have also been reported in MDD, which may be linked to HPA axis dysfunction [144], [145].

Clinical studies have shown that individuals who have been sexually or physically abused in childhood show an enhanced HPA axis activation both at baseline and when exposed to psychological stress [140], [146]. Early-life stress or trauma is a strong risk factor for adult depression, and a dose-response relationship between the severity of childhood adversity and the risk of lifetime and recent depressive episodes in adulthood has also been reported [115]–[117]. This link appears to be reflected in the altered ability of the HPA axis to respond to stress following childhood trauma, which can lead to an increased susceptibility to depression [147], [148]. HPA axis disturbances associated with childhood trauma include sensitisation of the neuroendocrine stress response, CRF hyperactivity, glucocorticoid resistance, and reduced hippocampal volume, which are comparable to many of the neuroendocrine abnormalities in depression [139], [146]. In addition, women show greater responsiveness to stress than men, consistent with the higher incidence of MDD in women [149].

1.2.5. Inflammation and immune response

The role of immune system dysfunction in the pathophysiology of major depression is supported by the observations that MDD patients exhibit activated inflammatory pathways [141]. Studies have reported that MDD patients show increased levels of pro-inflammatory cytokines, including interleukin (IL)-6, IL-1-β and tumour necrosis factor (TNF)-α, both in CSF and peripheral blood (serum or plasma) [150]–[156]. These findings have led to the cytokine hypothesis of depression. Acute-phase proteins, such as C-reactive protein (CRP), α-1-acid glycoprotein, α-1-antichymotrypsin and haptoglobin, as well as chemokines and cellular adhesion molecules, such as human macrophage chemoattractant protein-1, soluble intracellular adhesion molecule-1 and E-selectin, have also been found to be elevated in MDD

[141], [150], [151]. In addition, various inflammatory markers, including IL-6 and IL-1-β, have been reported to be positively correlated with depression symptom severity [153], [157], and administration of the inflammatory cytokine interferon-α has been shown to induce depressive symptoms [158].

Moreover, evidence suggests that immune/inflammatory responses may contribute to the development of MDD partially through interactions with the stress response pathways involving the HPA axis [141], [150], [159]. Given the communication that occurs between the immune and neuroendocrine systems, activation of immune/inflammatory responses could result in HPA axis hyperactivity, and vice versa [160]. As glucocorticoids that are released by the HPA axis in response to stress normally have anti-inflammatory and immunosuppressive effects, glucocorticoid resistance in MDD could induce immune activation; on the other hand, increased inflammation could lead to HPA axis hyperactivity by directly stimulating the CRF or by inducing glucocorticoid resistance [141], [159]. In addition, the relationship between early-life stress (childhood trauma) and MDD has been linked to not only HPA axis hyperactivity (as discussed above in **Section 1.2.4**), but also increased activation of the immune/inflammatory system [161], [162]. Stress and depression have been associated with increased levels of circulating cytokines and an increased activity of the brain immune cell microglia, indicating elevated inflammatory responses in the CNS as well as in the peripheral system [163], [164]. These findings suggest that inflammation and microglial activation play a role in the pathophysiology of depression. It has been proposed that the increase in cytokines following stressful life experiences in childhood may induce changes to microglia, which may, in turn, lead to structural and functional changes in the brain, most notably within the hippocampus and the prefrontal cortex, that increase the vulnerability to developing depression in adulthood [165], [166].

1.2.6. Gene-environment interactions

MDD is thought to result from a complex interplay of genetic predispositions and environmental exposures such as stressful life events (*i.e.*, gene-environment interaction). A gene-environment interaction is defined as a differential effect of an environmental exposure on disease risk in individuals with different genotypes; or alternatively, a differential effect of a genotype on disease risk in individuals with different environmental exposures [167]. Essentially, it reflects a mechanism whereby genetic and environmental factors jointly

contribute to the development of a disease, and the sensitivity to environmental exposures is influenced by genetic factors [168], [169]. Such interactions may help to account for the lack of consistent findings from GWAS of MDD [48], [135].

Although our understanding of how genetic and environmental factors interact in the pathogenesis of depression is incomplete, emerging evidence suggests that epigenetic mechanisms (such as deoxyribonucleic acid (DNA) methylation and histone modification), which modify gene function without changing the DNA sequence, may play a role in mediating the combined effects of genetic predisposition and environmental exposure on the risk for depression development [48], [105], [170]. Epigenetic modifications resulting from early-life trauma have been associated with an increased vulnerability to develop MDD in adulthood [168], [171], [172]. For example, the link between early-life stress, HPA axis dysfunction and increased risk of MDD in adults could possibly be explained by epigenetic regulation of the glucocorticoid receptor expression following childhood abuse [172], [173]. Therefore, genetic predispositions and stressful experiences in childhood may result in a maladaptive stress response system that amplifies the impact of negative life events and increases the vulnerability to depression in adulthood [168], [174]. In other words, whether or not individuals with adverse childhood experiences develop MDD (or other psychiatric disorders) in adulthood is dependent on the individual genetic background and its regulation of the stress response system [168].

1.2.7. Neurotrophic factors (neurogenesis and neural plasticity)

Observations of reduced volumes of the hippocampus and prefrontal cortex areas of the brain in depressed patients have led to a hypothesis of depression involving neurotrophic factors [175], [176]. Brain-derived neurotrophic factor (BDNF), a growth factor that regulates neuronal development and plasticity, is thought to play a particularly important role in depression [177]. Studies have found that while the expression of BDNF in the brain is reduced by stress, it is increased after treatment with antidepressants [175], [178]. In addition, reduced serum and plasma levels of BDNF have been reported in MDD patients [179]–[182].

1.2.8. Functional neural circuits

It has been suggested that stress-related alterations in inflammatory and glucocorticoid signalling could be associated with functional changes in corresponding brain networks [183], [184]. Consistent with this idea, neuroimaging studies of MDD have reported abnormalities in

the connectivity and/or activation of the affective-salience circuit, which plays an important role in guiding motivated behaviour, as well as the frontoparietal circuit, which is involved in cognitive control. Increased connectivity and activation of regions of the affective-salience circuit, including the amygdala, dorsal anterior cingulate and anterior insula, have been observed, which may reflect the heighted salience to negative stimuli in MDD [185]. Furthermore, MDD has been linked to reduced connectivity among regions of the frontoparietal network, particularly in response to negative stimuli; this may reflect general deficits in cognitive control of attention and emotion regulation in MDD, and implicate a role played by this network in inappropriate appraisals of negative life events [186], [187].

1.3. Blood-based proteomic biomarkers of MDD

1.3.1. The clinical need for biomarkers

Although extensive research has led to a better understanding of the pathophysiological mechanisms underlying MDD, there are currently no biological tests that have been approved and implemented into routine clinical practice for MDD or other psychiatric disorders [188]. Given the considerable disease burden and economic costs associated with undiagnosed or inadequately treated depression, there is a crucial yet unmet need to improve clinical management and patient care by developing more objective and robust methods of diagnosis and treatment selection. This can be achieved by identifying biomarkers of MDD, which would have valuable applications in disease detection and monitoring [189]. While biomarkers are routinely used to detect and monitor various medical conditions such as breast cancer, diabetes and heart disease, this is not yet the case for psychiatric disorders.

A biomarker, as defined by the National Institutes of Health Biomarker Working Group, is ‘a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention’ [189]. Biomarkers can be categorised into three types: diagnostic, prognostic, and treatment. Diagnostic biomarkers are useful in detecting the presence of a disease. They could facilitate earlier and more accurate diagnosis, and consequently enable patients to receive appropriate treatment at an earlier stage of the disease. Early diagnosis, by leading to early treatment, has been shown to beneficially affect the progression and outcome of MDD [190]. While clinical

diagnosis can only be established in the presence of sufficient symptoms (that fulfil the defined criteria), biomarkers may be detectable in patients in the months or even years preceding the overt manifestation of the disorder, which could be indicative of disease development [14]. Prognostic biomarkers characterise the course of the disease and predict disease outcome or severity (independent of treatment); they could be used to monitor disease progression [191]. Treatment biomarkers are useful in predicting treatment response. They could allow for the substratification of MDD patients based on their responses to different treatments, such as pharmacotherapy and psychotherapy, and guide personalised treatment strategies by helping clinicians to find an optimal approach for a particular patient with maximum benefit and minimum side effects [14], [42]. This is expected to reduce the duration of ineffective treatment period. Although biomarkers are typically considered to refer to explicitly biological measures, such as genetic or protein features, they are not necessarily limited to be so; features measured in surveys or rating scales could also represent candidate biomarkers [192]. Thus, biomarkers have the potential to facilitate earlier and more accurate diagnosis and treatment for patients, which would improve their quality of life, enable a better allocation of resources of the healthcare system, and reduce the disease-related burden and costs for both governments and patients [193]. In addition, biomarkers could help to elucidate the underlying biological and molecular mechanisms of disease.

1.3.2. Proteomics for biomarker discovery

Advances in '–omics' technologies have created novel opportunities for identifying molecular signatures, or biomarkers, of MDD. While genomics and transcriptomics were originally considered the state-of-the-art disciplines for biomarker discovery, with numerous reports pointing towards genetic and transcriptional aetiologies of psychiatric disorders, over the last decade, there has been a growing interest in proteomics, that is, the comprehensive analysis of all proteins in a biological system [194], [195]. As proteins are the functional molecules of the cell and reflect the dynamic state (current biological status) of the organism, a proteomic approach could enable the detection of disease-related alterations that are not detectable at the genetic or transcript level and help to elucidate the biological mechanisms underlying MDD [193], [196]. Recent developments in proteomic technologies mean that vast amounts of protein information, including protein expression levels, post-translational modifications, and protein-protein interactions, can be readily collected from bodily fluids and tissues [197], [198]. Hence, proteomics has emerged as a powerful approach for biomarker discovery and validation.

Proteomic studies have been performed using various methods, including two-dimensional gel electrophoresis, immunoaassays, protein microarrays, affinity separation, and mass spectrometry (MS)-based technologies. Immunoassays, such as enzyme-linked immunosorbent assays (ELISAs), are generally considered the gold standard in clinical laboratories. However, they have several drawbacks which can limit the technological and biological reproducibility of the results, including cross-reactivity, availability of suitable antibodies, batch-to-batch antibody variation, high cost, and requirement for relatively large sample volumes [199]–[202]. The limitations associated with immunoassays have encouraged a shift towards MS-based techniques in proteomic biomarker discovery and validation, as they allow high-throughput analyses with increased sensitivity and specificity and therefore yield more reproducible results [203]. This shift has been further facilitated by the fact that MS is already routinely used in clinical laboratories to detect and quantify metabolites, hormones, small molecules and drugs in blood, urine, and other biologically relevant material [204]–[206].

1.3.3. Mass spectrometry

MS enables the identification and quantification of proteomic analytes by measuring their mass-to-charge ratio *(m/z)* [207]. Given its high sensitivity and multiplexing capacity to simultaneously characterise and quantify several thousands of analytes, it is emerging as an increasingly powerful method for proteomic biomarker research [208]–[210]. In particular, MS coupled with separation techniques such as liquid chromatography (LC-MS) has become an effective tool for providing a comprehensive analysis of the proteome and profiling of disease-related alterations [211], [212]; changes in protein expression, abundance, structure, and function, all of which could serve as potential indicators of disease, can be detected [213]. Thus, the application of MS-based methods for molecular profiling of MDD has the potential to improve our understanding of the involved pathophysiological mechanisms as well as identify diagnostic, prognostic and treatment biomarkers [194].

Proteomic analysis by MS can involve a targeted or an untargeted approach. Untargeted analysis (*e.g.*, shotgun proteomics) aims to identify and quantify all detectable proteins or analytes present in the sample (global profiling). However, it can have limited sensitivity and precision, especially in the analysis of lower-abundance proteins [209]. On the other hand, targeted MS techniques such as multiple reaction monitoring (MRM) enable the identification

and quantification of a pre-selected set of proteins with increased sensitivity, precision, and throughput [210]. While untargeted methods are typically employed in biomarker discovery experiments to give a global characterisation of proteomic signatures, it is often desirable to subsequently confirm the findings using more reliable targeted methods [214], [215]. Thus, targeted and untargeted MS techniques are complementary for biomarker discovery and validation.

1.3.4. Biomarkers in peripheral blood

In order to be useful in routine clinical practice, biomarkers need to not only have high sensitivity and specificity (> 80%), but also be reliable, reproducible, non-invasive and cost-effective [216]. While the search for biomarkers of neuropsychiatric disorders has traditionally focused on the CNS (post-mortem brain tissue and CSF), there is converging evidence that disease-related alterations can also be detected in the peripheral system, such as the circulating blood [14]. Blood is a widely used sample type for biomarker discovery, as it is a rich source of proteins that reflects the complexity of the human proteome as well as a circulating representation of the physiological and pathophysiological processes occurring in the body [217]–[219]. Extensive research has shown that depression is associated with changes and/or disturbances in the function of several biological systems, which indicate that depression is a systematic disorder. Consistent with this, proteomic studies of MDD have reported alterations in peripheral biomarkers involved in the HPA axis, immune/inflammatory response, as well as carbohydrate and/or lipid metabolism [14], [130], [220]–[227]. The accessibility, low invasiveness, time-efficiency and cost-effectiveness of blood-based peripheral biomarkers make them an appealing alternative to CNS biomarkers for clinical application with regards to the detection and monitoring of psychiatric disorders [14], [228], [229]. The utility of blood-based biomarkers has already been demonstrated in other clinical fields. This is especially the case for oncology, as various blood-based protein biomarkers have obtained FDA clearance and are currently used in clinical practice for cancer screening, monitoring, prognosis and/or prediction. These include: cancer antigen 125 for ovarian cancer; cancer antigen 19-9 for pancreatic cancer; carcinoembryonic antigen for colon cancer, prostate-specific antigen for prostate cancer; and cancer antigen 15-3, oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 for breast cancer [230]–[233]. Additionally, a diagnostic blood test for liver fibrosis, which measures three serum biomarkers, hyaluronic acid, procollagen III amino terminal peptide, tissue inhibitor of metalloproteinase 1, has been

developed and made available for clinical use [234]. This field is evidently lagging behind in psychiatry as a blood-based biomarker test is yet to be approved by the FDA or other regulatory authorities and successfully make its way into clinical application [235].

Given the complexity and heterogeneity of MDD, single biomarkers are unlikely to sufficiently capture the disease characteristics. Hence, it is anticipated that a panel of multiple biomarkers will be needed to not only provide appropriate coverage of the various biological abnormalities, but also achieve sufficient sensitivity and specificity, that is, deliver sufficiently high discriminatory power, for clinical implementation [14], [223], [236], [237]. For instance, it has been shown that although serum proteins have between small and moderate effect sizes when considered individually, they are able to distinguish between MDD patients and controls with a good discriminatory power when combined into a biomarker panel [222]. Additionally, as biomarkers involved in disease-related pathways tend to be functionally highly correlated, it is important that the joint effects of the biomarkers are investigated using appropriate statistical methods [222].

1.3.5. Blood-based diagnostic biomarkers of MDD

Recent studies have demonstrated that blood proteomic profiling enables the identification of differentially expressed proteins between individuals with and without MDD as potential diagnostic biomarkers of MDD. To provide a comprehensive overview of the current evidence for blood-based diagnostic biomarkers of MDD, a PubMed database search was conducted using the search terms: ‘depression proteomic biomarker blood’, ‘depression proteomic biomarker serum’ and ‘depression proteomic biomarker plasma’ (*Preece, Han, and Bahn (2018)* [228]). The search was based on articles available as of September 2020, and was restricted to human studies published in the last ten years. The ten-year period was chosen to ensure that similar sample formats and analytical techniques were implemented in the studies to allow for a straightforward comparison, given the increase in the number of studies using peripheral blood as a sample source and the notable developments in proteomic technologies. Biomarkers associated with predicting disease prognosis or treatment response were not considered. The PubMed database search was conducted and the findings were analysed and summarised in collaboration with Rhian Lauren Preece, who was another candidate PhD student in the Bahn Lab.

This search strategy identified 11 studies focused on identifying diagnostic biomarkers that differentiate between individuals with and without MDD [222], [223], [244], [226], [227], [238]–[243]. These studies are summarised in **Table 1.2**. Two studies used plasma [227], [238] and nine studies used serum [222], [223], [226], [239]–[244] as the sample source of peripheral blood. The studies show that a range of proteomic techniques have been applied for protein biomarker discovery. MS- and immunoassay-based techniques have been used either alone or in combination, where immunoassays have been used to validate findings from MS analyses. Studies using single or multiplex immunoassay profiling have adopted a targeted approach and measured candidate protein biomarkers that have been identified as proteins of interest in previous studies and/or neurobiological theories of affective disorders. In contrast, other studies have adopted an untargeted approach (*e.g.*, LC-MS/MS) to identify potential blood protein biomarkers. In terms of the statistical methods, univariate methods (Student's t-test, analysis of variance (ANOVA), analysis of covariance (ANCOVA), and Mann-Whitney *U* test) have been predominantly used, where proteins have been reported as significantly altered between MDD patients and controls if *p*-values were below a pre-defined threshold (often 0.05; before or after multiple testing correction). Some studies have implemented one or more multivariable methods with variable selection (logistic regression, least absolute shrinkage and selection operator (LASSO), random forests, discriminant analysis, principal component analysis (PCA), or support vector machine) to identify differentially expressed protein biomarkers.

The studies demonstrate that the comparison of blood (serum or plasma) proteomic profiles enables the identification of proteins that are differentially expressed between MDD patients and healthy controls. These proteins have functional roles in immune response, inflammation, metabolism, and cell signaling, among others. Across the 11 studies, 84 unique proteins were identified as diagnostic biomarker candidates for MDD. Ceruloplasmin was reported in five studies [226], [238], [240], [241], [243], and seven proteins (α-2-macroglobulin [222], [227], [238], apolipoprotein B-100 [238], [241], [243], apolipoprotein D [238], [239], [241], BDNF [223], [226], [227], IL-1 receptor antagonist protein [222], [226], [239], macrophage migration inhibitor factor [222], [226], [239], and protein S100-A12 [222], [226], [239]) were each reported in three studies. While proteins that have been repeatedly reported across different studies may potentially represent diagnostic biomarkers of MDD and should be further investigated, there were disagreements in the direction of change in some cases. Therefore, care must be taken when comparing and interpreting findings from different studies, as they

vary on the proteomic technique used, the proteins quantified, and the statistical method applied [228].

Of paricular note is the work by *Papakostas et al. (2013)* [223] and subsequently *Bibello et al. (2015)* [245], which represented one of the first attempts in the field to develop a panel consisting of multiple biomarkers and to explore a probabilistic approach. This eventually led to the commercialisation of a diagnostic test for MDD (MDDScore) by Ridge Diagnostics Inc based on a panel of nine serum biomarkers, whereby the probability of an individual having depression could be calculated using the biomarker concentrations and body mass index (BMI). The panel was developed in a pilot study of 36 MDD patients and 43 non-depressed controls and a replication study of 34 MDD patients [223], and later validated in a larger follow-up confirmation study of 68 MDD patients and 86 non-depressed controls, which achieved sensitivity and specificity above 90% and 80%, respectively [245]. The nine serum biomarkers (α-1 antitrypsin, apolipoprotein CIII, BDNF, cortisol, epidermal growth factor, myeloperoxidase, prolactin, resistin, and soluble TNF-α receptor type II) are associated with inflammatory, HPA axis, metabolic and neurotrophic pathways. However, despite the high accuracy, a potential limitation of this test is that it is not known to have been tested on naturalistic patient populations recruited from the general population and primary healthcare settings, where the need for a diagnostic test for MDD is greatest (the published studies were conducted in more restricted patient populations). Additionally, it has not yet been approved by official regulatory authorities such as the FDA.

Moreover, *Bot et al. (2015)* [239] previously conducted the largest proteomic study examining serum proteins in 687 currently depressed patients (six-month recency of depressive episode) and 420 controls recruited from different healthcare settings (general population, primary care and specialised mental healthcare). Thirty-three analytes were identified as significantly different in MDD patients compared to controls based on univariate analysis, where p-values < 0.05 were considered as statistically significant. However, this study did not consider the joint effects of the biomarker candidates, which are essential for exploring their diagnostic potential.

More recently, *Chan et al. (2016)* [222] performed a meta-analysis of multiplex immunoassay profiling data to develop and validate a diagnostic biomarker panel for detecting depression. A panel of 33 serum proteomic biomarkers, which defined an immune-neuroendocrine profile in

depression, was identified by applying a joint effects statistical method (LASSO regression with ten-fold cross-validation) to differentiate between 78 first-/recent-onset drug-naive/drug-free MDD patients and 156 controls. Importantly, the discriminatory performance of the panel was subsequently tested on 468 currently depressed patients (one-month recency of MDD diagnosis) and 305 controls recruited from naturalistic healthcare settings (general population, primary care and specialised mental healthcare). This resulted in a moderate discriminatory performance, which further improved following the incorporation of sociodemographic variables, such as age, BMI, family history and alcohol dependence and/or abuse. This highlighted the importance of considering sociodemographic variables in addition to the proteomic biomarkers for differentiating between patients and controls. Despite the need for further validation, the authors demonstrated the feasibility of a blood-based biomarker-sociodemographic panel for detecting depression and improving the accuracy of clinical diagnosis in naturalistic healthcare settings [222].

One of the challenges in identifying robust diagnostic biomarkers of MDD arises from the fact that the pathophysiology of MDD overlaps, at least to a certain extent, with that of other major psychiatric disorders, including BD and schizophrenia. GWAS have demonstrated that genetic risk factors are shared between MDD, BD and schizophrenia [246]–[248], and disturbances in the immune/inflammatory and neuroendocrine pathways have been implicated across these disorders [150], [249]–[251], suggesting a potential overlap in biomarker profiles between MDD and other major psychiatric disorders. In line with this, a recent review of previous blood-based biomarker studies has demonstrated that the same proteins (such as TNF-α, CRP and interferon-γ) have been identified as biomarkers of not only MDD but also BD and/or schizophrenia [14]. Similarly, another review, *Preece, Han, and Bahn (2018)* [228], has shown that, among 82 proteins which have been identified as biomarkers of MDD, 29 (including apolipoprotein AI, angiotensin-converting enzyme and CD40 ligand) have also been identified as biomarkers of BD. These proteins may be involved in disease-related alterations and pathophysiological mechanisms that are common to both MDD and BD, and could reflect risk factors that are shared between the two disorders [252]. Thus, this complication that biomarkers may not be disease-specific reinforces the need for a diagnostic test for MDD to be based on a panel of multiple biomarkers, as a test based on any single biomarker is likely to lack sufficient discriminatory power [14], [193]. Furthermore, considering the joint effects of biomarkers and other patient characteristics, including sociodemographic factors and/or clinical symptoms, will be important, as this approach may reveal a disease-specific profile of MDD that consists

of a combination of biological and non-biological features. The idea is that, based on such a profile, an individual presenting with only the biological features of MDD (*e.g.*, an inflammatory profile) may be predicted as having a relatively low probability of having or developing MDD in comparison to another individual presenting with both the biological and non-biological features of MDD.

Table 1.2. Proteomic studies identifying blood-based diagnostic protein biomarkers of MDD (published in the last ten years).

The sample groups, biological material, proteomic technique, statistical method, number of proteins or analytes targeted, and the findings, are shown for each study. *Proteins that are reported in two studies, independent of the direction of change. **Proteins that are reported in three or more studies, independent of the direction of change. Abbreviations: ANCOVA (analysis of covariance); ANOVA (analysis of variance); ELISA (enzyme-linked immunosorbent assay); iTRAQ (isobaric tag for relative and absolute quantification); LASSO (least absolute shrinkage and selection operator); LC (liquid chromatography); MALDI-TOF (matrix assisted laser desorption ionisation time of flight); MS (mass spectrometry); MDD (major depressive disorder); PCA (principal component analysis); PLS-DA (partial least squares discriminant analysis); RFE (recursive feature elimination); SVM (support vector machine).

Study	Material	Sample groups	Proteomic technique	Statistical method	Number of proteins or analytes targeted	Findings	Upregulated	Downregulated
Domenici et al. (2010) [227]	Plasma	MDD patients (n=245) *vs* healthy controls (n=254)	Multiplex immunoassay	Student's t-test; ANOVA; PCA; PLS-DA; Random Forests	79 proteins	2 proteins differentially expressed (univariate); 7 important proteins identified by multivariable analysis (high variable importance of contribution from PLS-DA)	Insulin*; Matrix metalloproteinase-9; Plasminogen activator inhibitor 1; Tumor necrosis factor receptor superfamily member 1B*	Brain-derived neurotrophic factor**; Apolipoprotein A-I; α-2-macroglobulin**
Xu et al. (2012) [238]	Plasma	First-onset, treatment-naïve depressed patients (n=21)	iTRAQ LC-MS/MS; validation by immunoblotting, ELISA	Mann-Whitney *U* test	Untargeted (LC-MS/MS)	9 proteins differentially expressed	Apolipoprotein D**; Afamin; Apolipoprotein B-100**; α-1B-glycoprotein;	Ceruloplasmin**; Histidine-rich glycoprotein; Semaphorin-3F;

		vs healthy controls (n=21)					Vitamin D-binding protein isoform 1	α-2-macroglobulin**
Papakostas et al. (2013) [223]	Serum	MDD patients (pilot n=36, replication n=34) *vs* healthy (n=43)	ELISA (individual immunoassays)	ANCOVA	110 proteins	9 proteins differentially expressed	α-1-antitrypsin*; Brain-derived neurotrophic factor**; Cortisol*; Pro-epidermal growth factor; Myeloperoxidase; Tumor necrosis factor	Apolipoprotein C-III; Prolactin; Resistin
Stelzhammer et al. (2014) [226]	Serum	First onset, drug-naïve MDD patients (discovery n=23, validation n=15) *vs* healthy controls (discovery n=42, validation n=21)	Multiplex immunoassay; LC-MS^{E}	Mann-Whitney *U* test; ANCOVA	190 analytes (multiplex immunoassay); Untargeted (LC-MS^{E})	11 proteins differentially expressed (multiplex immunoassay); 2 proteins differentially expressed (LC-MS^{E})	Ferritin*; Macrophage migration inhibitory factor**; Protein S100-A12**; Superoxide dismutase; Interleukin-1 receptor antagonist protein**; Pro-interleukin-16; Tenascin; Brain-derived neurotrophic factor**; Haptoglobin-related protein	Angiotensin-converting enzyme*; Serotransferrin; Somatotropin*; Ceruloplasmin**
Bot et al. (2015) [239]	Serum	MDD patients (current n=687) *vs*	Flow cytometry (multi-analyte profiling)	Univariate linear	171 proteins	28 proteins differentially expressed	Pancreatic prohormone*; Prostasin;	Follicle-stimulating hormone*;

	healthy controls (n=420); validation on antidepressant-free current MDD patients (n=78) *vs* healthy controls (n=156)	regression	Angiogenin; Apolipoprotein D**; α-1-antitrypsin*; α-1-antichymotrypsin; Macrophage migration inhibitory factor**; Urokinase plasminogen activator surface receptor; Stromelysin-1*; Cathepsin D; Serine protease hepsin; Stromelysin-2; Protein S100-A12**; Interleukin-12 subunit β*; Interleukin-1 receptor antagonist protein**; Apolipoprotein A-IV; Complement factor H-related protein 1; α-2-HS-glycoprotein; Carcinoembryonic antigen; von Willebrand factor*; Fatty	Luteinizing hormone; Growth-regulated α protein*; Insulin-like growth factor-binding protein 5; Angiopoietin-2; Receptor tyrosine-protein kinase erbB-3; CD40 ligand

							acid-binding protein, adipocyte	
Lee et al. (2015) [240]	Serum	MDD patients (n=8) *vs* healthy controls (n=8)	ELISA	Mann-Whitney *U* test	10 proteins	3 proteins differentially expressed	Ceruloplasmin**; Complement C1q subcomponent subunit C; Inter-α-trypsin inhibitor heavy chain H4*	
Lee et al. (2016) [241]	Serum	Drug-free female MDD patients (LC-MS/MS n=10, MRM n=25) *vs* healthy controls (LC-MS/MS n=10, MRM n=25)	LC-MS/MS; MRM	LASSO	Untargeted (LC-MS/MS)	6 proteins differentially expressed	Apolipoprotein D**; Apolipoprotein B-100**; Vitamin D-binding protein*; Hornerin	Profilin-1; Ceruloplasmin**
Wang et al. (2016) [242]	Serum	First-onset treatment-naïve MDD patients (n=22) *vs* healthy controls (n=20)	iTRAQ LC-MS/MS; validation by ELISA	Student's t-test	Untargeted (LC-MS/MS)	4 proteins differentially expressed	C-reactive protein; Angiopoietin-related protein 3; Serum amyloid A-1 protein; Inter-α-trypsin inhibitor heavy chain H4*	
Chan et al. (2016) [222]	Serum	First-onset drug-naïve drug-free MDD patients (n=78) *vs* healthy controls (n=156)	Multiplex immunoassay	LASSO; Stepwise logistic regression	99 analytes	33 proteins differentially expressed	α-2-macroglobulin**; von Willebrand factor*; Serum glutamic oxaloacetic transaminase; Insulin-like growth factor-binding protein 2;	Factor VII; Angiotensin-converting enzyme*; Creatine kinase MB; Tyrosine-protein kinase receptor UFO; Stromelysin-1*; Thyroxine-

							Follicle-stimulating hormone*; Insulin*; Pancreatic prohormone*; Angiotensinogen; Cortisol*; Interleukin-1 receptor antagonist protein**; Interleukin-1 α; C-C motif chemokine 22; Macrophage migration inhibitory factor**; Protein S100-A12**; Ferritin*; Cancer antigen 19-9 (CA 19-9); Tumor necrosis factor receptor superfamily member 10C	binding globulin; Thyroid-stimulating hormone; Somatotropin*; Interleukin-13; Granulocyte colony-stimulating factor; Growth-regulated α protein*; C-C motif chemokine 2; Tumor necrosis factor receptor superfamily member 1B*; Interleukin-12 subunit β*; Interleukin-3; Myoglobin
Kim et al. (2017) [243]	Serum	Drug-free female MDD patients (n=25) *vs* healthy controls (n=25)	LC-MS/MS; MRM	SVM-RFE	Untargeted (LC-MS/MS)	3 proteins differentially expressed	Apolipoprotein B-100**; Vitamin D-binding protein*	Ceruloplasmin**

Kang et al. (2019) [244]	Serum	MDD patients (n=86) *vs* healthy controls (n=89)	MALDI-TOF MS	SVM	Untargeted	2 proteins differentially expressed	Fibrinogen	Kininogen

1.4. Statistical methods for biomarker identification

1.4.1. Clinical risk prediction models

Clinical risk prediction models estimate the probabilities or risks of having (diagnostic) or developing (prognostic) an outcome or a disease [253]. They are a useful tool to integrate various patient characteristics, including molecular profiles, sociodemographic factors and symptoms, to make probabilistic predictions in a multivariable manner [254], [255]. These prediction models can be used within a clinical setting to facilitate, rather than replace, decision-making by healthcare providers, as the probabilities estimated by the models can aid in making more statistically informed and objective decisions on the diagnosis, prognosis, and treatment of patients [256], [257]. This supports the idea of personalised medicine and would increase the efficiency and cost-effectiveness of clinical management [258].

Developing a prediction model involves two stages: model development and model evaluation (**Figure 1.1**). In the development stage, multivariable classification and model selection methods are used to fit a prediction model to the training data. Any patient features that are known or considered to be associated with the outcome, such as sociodemographic factors and candidate biomarkers, can be investigated as a predictor [254]. Multivariable methods select a set of predictors with the best joint ability for discrimination, and define the individual contribution of the selected predictors to the probability estimation of the outcome (in the form of regression coefficients) [254]. The probability of having or developing the outcome can then be calculated for each individual based on the combined contributions of the selected predictors.

In the evaluation stage, the predictive ability of the fitted prediction model is assessed when applied to data other than that used for model development (test data). The selected model is used to predict the outcomes of a new set of individuals and the predictions are compared against the observed outcomes [253]. This allows for an evaluation of the generalisation performance of the prediction model, which is highly important for clinical application [259]. A diagnostic prediction model should be able to accurately and reliably distinguish between individuals with and without disease (discrimination). Ideally, independent participant datasets should be used for model development (training data) and model evaluation (test data) if available, and both training and test datasets should be sufficiently large and diverse to ensure

that they are representative, although such data-rich situations tend to be rare in clinical studies [260].

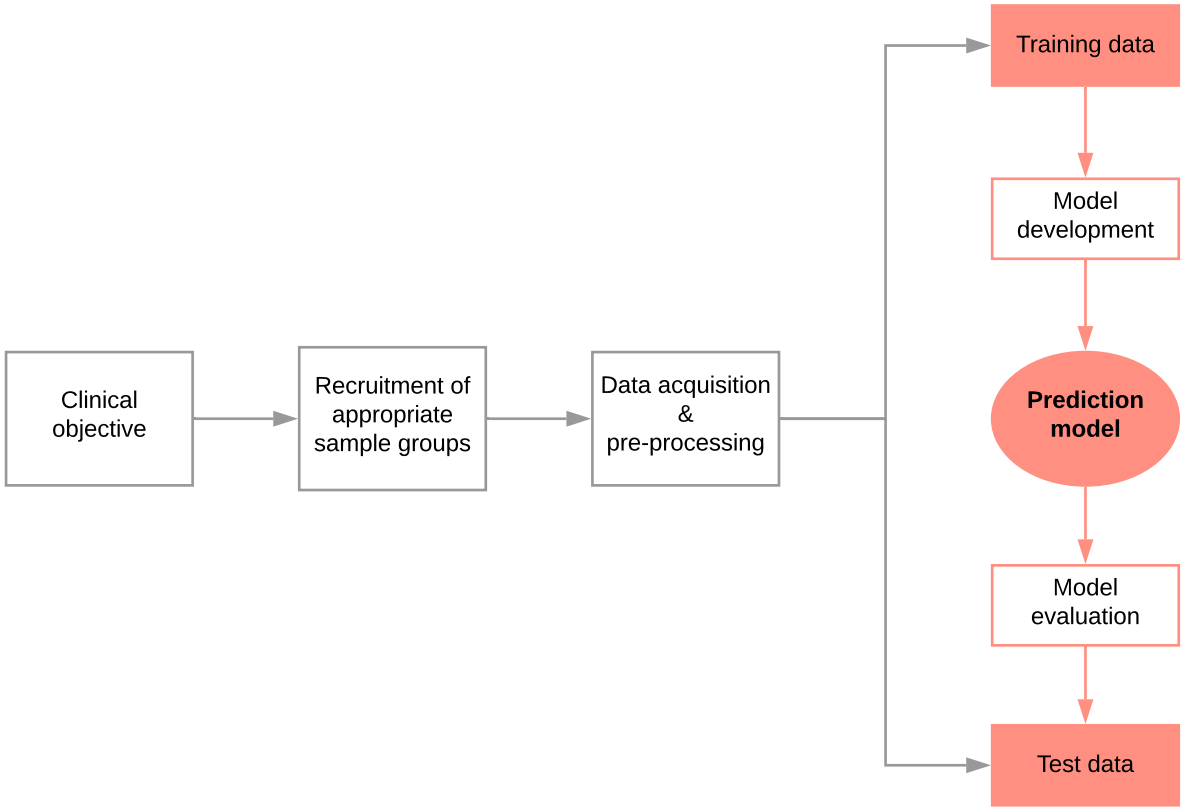


Figure 1.1. Workflow of prediction model development and evaluation.

1.4.2. Statistical considerations for biomarker discovery and validation

Despite considerable research and investment into proteomic biomarker discovery for MDD and other psychiatric disorders, very few biomarkers have been successfully translated into regulatory authority-approved clinical tests and are routinely used in clinical practice [261], [262]. This is mainly because many candidate biomarkers that were initially found to be discriminatory have failed to be validated in subsequent studies [263], [264]. As in other fields of biomedical research, the lack of reproducibility of potential biomarker findings for psychiatric disorders arises from not only the inherent complexity and heterogeneity of the disorders, but also limitations in the pre-analytical, analytical and statistical workflow [15], [261], [265]. This suggests that the task of developing and validating a biomarker panel with a sufficiently high discriminatory power may be more challenging than previously considered.

1.4.2.1. Study design

While appropriate study design is an important pre-analytical factor that determines the reproducibility of findings, a major limitation of many biomarker studies is the fact that they are not designed with a specific clinical objective and therefore are carried out without strict recruitment criteria for sample inclusion or exclusion [15]. For the identification of diagnostic biomarkers, first- or recent-onset drug-naïve patients should ideally be compared against controls. Using established or chronically ill patients or patients on medication is less ideal not only as the effect of medication can be a confounding factor in the analysis, but also as the findings may have limited generalisability when applied for the diagnosis of first- or recent-onset patients [15]. In addition, rather than using 'healthy' controls as a reference population for comparison against patients, using individuals presenting with subthreshold depression may allow for a more clinically relevant and appropriate reference sample population, as distinguishing MDD patients from subthreshold symptomatic individuals is more likely to represent a task that GPs would need to undertake in primary care compared with distinguishing MDD patients from non-symptomatic individuals. Furthermore, the presence of comorbid diseases (*e.g.*, MDD with anxiety disorder) or symptoms of other undiagnosed diseases (*e.g.*, MDD with subthreshold symptoms of anxiety) increases the heterogeneity of the population, making discrimination more difficult. While this problem could be addressed by using a more homogeneous sample population, it may not be truly representative of the target population, and findings would be less generalisable when applied to more naturalistic patient cohorts. Since studies that lack proper design tend to be carried out using pre-existing or available samples and data that have originally been collected for other purposes, it is expectedly less likely to be able to obtain reproducible findings from such studies [15]. The poor reproducibility of findings is also a result of many biomarker studies being conducted using small sample sizes [261] (see **Section 1.4.2.2** below).

Moreover, as individuals' blood-based biomarker profiles can vary substantially with sociodemographic and lifestyle factors, including age, sex, BMI, smoking and alcohol consumption, these variables could be potential confounders [15]. A confounder is a variable that is related to both the predictor variables and the outcome variable, and results in a spurious association between the predictors and the outcome [266]. It is therefore important that potential confounders are appropriately dealt with, for instance, as part of the study design by recruiting controls that are matched with patients for certain sociodemographic and lifestyle factors. However, as this can be challenging to achieve and/or result in non-matched patients

or controls having to be excluded (*i.e.*, wasteful of resources), another approach is to adopt appropriate statistical techniques to account and adjust for the effects of potential confounders [267]. For example, multivariable regression methods, such as logistic regression, investigate the joint effects of many variables, including predictors and potential confounders, on the outcome, and thereby enable potential confounders to be dealt with in a relatively flexible manner [266]. Additionally, reproducibility can be affected by technical variability, including variability in sample collection, storage, and handling [268]. Establishing strict guidelines (standard operating procedures) which all participating centers need to follow would help to address these factors.

1.4.2.2. Model overfitting

Prediction model development in biomarker studies typically involves datasets that contain a large number of predictors (p) relative to the number of observations (n). Performing model selection on such high-dimensional datasets ('small n, large p') is likely to lead to model overfitting, which is when the prediction model captures not only the underlying relationship of interest but also noise in the data [255]. As a result, the model provides an over-optimistic assessment of the predictive performance based on the original data, which does not reproduce when the model is applied to new data. An overfitted model typically underestimates the probability of an event in low risk individuals and overestimates it in high risk individuals [269].

When developing a risk prediction model, the events per variable (EPV), which is the number of observations (in the smallest sample group) divided by the number of candidate predictors, is often used for sample size considerations. Candidate predictors should include all variables that are initially considered in the study as potential predictors and not only those used in multivariable analysis for prediction model development, as well as dummy variables that are used for categorical predictors [270]. As a rule-of-thumb, it has been recommended that an EPV of at least ten is required to avoid overfitting and ensure that the prediction model is reliable; lower EPV values in prediction model development have been found to result in poorer predictive performance in validation [271]–[273]. Although this criterion has been questioned for not being based on convincing scientific reasoning [274], [275] and challenged for being too lenient [276], [277] as well as too strict [273], the EPV is nonetheless generally regarded as useful for considering the sample size in relation to model complexity and hence frequently reported in clinical prediction studies [270]. However, the recommended EPV ≥ ten

is often difficult to achieve in practice, especially for studies of psychiatric disorders where the recruitment of relevant patients and control individuals is notoriously challenging [193], [228]. Hence, appropriate statistical methods that can address the problem of overfitting need to be considered in order to ensure that a prediction model which only captures the meaningful and relevant aspects of the data, and thereby, is generalisable and useful for clinical decision-making, is obtained. A review of flawed research (by the Institute of Medicine at Duke University, USA) underscored the serious repercussions of model overfitting and incorrectly used statistical methods in –omics biomarker discovery studies, which subsequently led to not only the cancellation of clinical trials but also the retraction of many manuscripts [278].

Data prefiltering (dimensionality reduction)

A common approach that is used in proteomic biomarker discovery workflows to deal with the 'small *n*, large *p*' problem is to reduce the dimensionality of the dataset (and thereby increase the EPV) by prefiltering the predictor variables. The relationship of each predictor to the outcome is assessed based on univariable screening (such as Student's t-test, ANCOVA and Mann-Whitney *U* test), and only predictors that are statistically significant are used for subsequent model selection, whereas others are eliminated. However, when the same data that is used to prefilter the predictors is again used to build a prediction model, the generalisability of the model can be limited, as prefiltering gives an over-optimistic assessment of the predictive ability of the model [261]. Hence, this strategy is not recommended [279].

Regularisation

Alternatively, regularisation (or shrinkage) methods can be used to alleviate overfitting. Penalised regression methods, such as ridge regression and LASSO, are recommended in the Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) Statement for developing and validating prediction models [253], and are being increasingly used to build clinical risk prediction models with high-dimensional data [279]. By imposing a constraint on the values of the regression coefficients, these methods shrink (or regularise) the regression coefficients and consequently reduce the range of the predicted probabilities [255]. Ridge regression shrinks the regression coefficients towards zero by placing a constraint on the sum of the squared regression coefficients [280]. On the other hand, as LASSO constrains the absolute values of the regression coefficients, it shrinks the regression coefficients exactly to zero, which effectively eliminates poor predictors from the model [281]. Therefore, the LASSO method might be preferred to obtain a prediction model for potential

use in a clinical setting, as it yields a simpler model using fewer predictors without affecting the predictive ability of the model, which can save time and/or resources by allowing for less information to be collected from patients and help-seekers [255].

Validation

Ideally, the prediction model should be validated on independent patient data (external validation) to obtain an unbiased estimate of model predictive performance and evaluate whether the model is overfitted. When this is not available, internal validation techniques that use only the original data, such as cross-validation and bootstrapping, should be implemented as part of model development to limit overfitting and ensure that the obtained model is reliable [282].

1.4.2.3. Model selection uncertainty

Another problem that can limit the reproducibility of a prediction model is model selection uncertainty, which is when no single model is strongly supported by the given data. Traditionally, the process of model selection is concerned with selecting, from a set of all possible candidate models, a single model that best approximates the given data [283], [284]. Once the 'best' model has been determined, it is often treated as if it were the only model explaining the data, and subsequent inferences and predictions are based on this model alone. However, this approach ignores any uncertainty in model selection, which is inherently involved in the process of searching through the set of candidate models to identify the best one [285]. Model selection uncertainty can, for instance, arise from small changes in the data and be problematic as it reflects statistical variation that is not captured within the single 'best' model [285]. Failure to account for this uncertainty may lead to biased parameter estimates and inferences and predictions that are overconfident and not reproducible [286], [287]. Although model selection uncertainty often outweighs other sources of uncertainty (such as sampling variation and parameter uncertainty), it is typically overlooked in practice [288].

In cases where the 'best' model can clearly be identified, then inferences and predictions can be based on that model alone [284]. However, when model selection uncertainty is evident, it would be disingenuous to only present the 'best' model. Therefore, in these cases, it would be more appropriate to adopt a multimodel approach and derive inferences and predictions from a candidate set of models rather than from a single model [284], [289], [290].

1.4.3. Other challenges in the translation of biomarkers into clinical practice

Despite substantial advances, the progression from biomarker discovery to validation and clinical implementation remains a significant hurdle and requires cooperation and collaboration between many stakeholders. In addition to the abovementioned limitations associated with the lack of reproducibility of biomarker findings, there are challenges arising from funding and regulation. The financial burden increases as candidate biomarkers progress through the development pipeline (*i.e.*, towards the clinic) and validation studies need to be conducted with increasingly large sample sizes, which often results in academic investigators seeking financial support from the industry [233]. Yet, in spite of the potential clinical utility of a candidate biomarker, acquiring an industrial partner to undertake its development may be difficult if it is not considered to have sufficient commercial value and potential for investment return [15], [233]. In line with this, the inadequate reimbursement rates and the accompanying high financial risks of diagnostics relative to therapeutics have limited investments into the development and commercialisation of diagnostic tests [291]. Hence, academic investigators are advised to apply for intellectual property rights early in the biomarker discovery process to incentivise investment. Moreover, the translation of biomarkers into clinical practice is complicated by the stringent FDA regulations, under which protein-based (and other molecular) biomarker tests used for disease diagnosis, monitoring and treatment are considered as 'medical devices' and follow the same regulatory standards as other types of medical devices [292]. It is therefore important to ensure that the translational scientists working on biomarker development have a comprehensive understanding of not only the analytical and clinical requirements but also the regulatory requirements for biomarker tests [265], [292]. Additionally, as obtaining regulatory approval does not automatically lead to acceptance by the clinical community, demonstration of strong clinical utility, including the benefit that will be brought to the patients as well as feasibility in the clinic, is essential [293].

1.5. Thesis aims and outline

With less than half of patients with MDD correctly diagnosed within the primary care setting, there is a clear clinical need to develop an objective and readily accessible test to aid in earlier and more accurate diagnosis of the condition. Blood-based proteomic biomarkers, in combination with sociodemographic and other patient characteristics, can be used to build risk prediction models that could be implemented within a clinical setting to complement and facilitate evaluations by healthcare professionals and thereby improve patient care. The main aim of this thesis is to explore the potential of such clinical risk prediction models to provide a novel and robust diagnostic tool for MDD. This will involve:

i) Identifying blood-based proteomic biomarkers and sociodemographic and other patient characteristics as predictors of MDD, and
ii) Achieving reproducible predictions of the risk of having or developing MDD in clinically relevant individuals.

The remaining chapters of this thesis will be organised as follows:

- **Chapter 2** discusses the experimental and statistical methods used to obtain the results presented in subsequent chapters of the thesis. This includes an outline of the targeted MS technique used for serum and dried blood spot (DBS) proteomic profiling. Statistical data pre-processing and analysis methods are also described.

- **Chapter 3** explores the development of a novel approach which combines feature extraction and model averaging to account for model selection uncertainty as well as limit model overfitting, and thereby improve model predictive performance and reproducibility. The utility of this multimodel approach in predicting the probability of having a complex psychiatric disorder is investigated, and a dataset consisting of first-onset drug-naïve schizophrenia patients and controls is used to demonstrate proof-of-concept. The applications of this method in developing robust clinical prediction models of MDD are explored in **Chapters 4 and 5**.

- **Chapter 4** investigates the prediction of future onset of depression in individuals presenting with subthreshold levels of depressive symptoms using their serum proteomic, sociodemographic and clinical data. The multimodel approach combining feature extraction and model averaging, developed and refined in **Chapter 3**, is applied to develop a parsimonious prediction model of MDD in the presence of model selection uncertainty. The feasibility of detecting early disease indications in individuals with subthreshold depression prior to clinical diagnosis is explored.

- **Chapter 5** explores the identification of MDD among individuals presenting with subthreshold depression using data from DBS proteomics and a novel digital mental health assessment. The multimodel approach is applied to develop prediction models of MDD, and a repeated nested cross-validation approach is used to evaluate and address the variation in feature selection and predictive performance that arises from choosing different splits of the data. The ability of the prediction models to identify currently depressed patients without an existing MDD diagnosis as well as currently not depressed patients with an existing MDD diagnosis among individuals with subthreshold depression is assessed.

- Finally, **Chapter 6** provides an integrated summary of the findings from **Chapters 3, 4 and 5** and discusses their significance and clinical implications. The limitations of the work in this thesis and considerations of future studies in relation to the presented findings are also discussed.

Note how each publication relates to each chapter in the thesis:

- *Preece, Han, and Bahn (2018)* [228] relates to **Chapters 1 and 2**.

- *Cooper et al. (2019)* [294] relates to the work in **Chapter 3**.

- *Han et al. (2019)* [295] relates to the work in **Chapter 4**.

- *Olmert et al. (2020)* [296] and *Han et al. (2020)* [297] relate to the work in **Chapter 5**.

Chapter 2 Methods

This chapter outlines the experimental and statistical methods used to obtain the results presented in subsequent chapters of the thesis.

2.1. Participants

Participants were selected to comprise patient and reference groups in relevance to the clinical objectives of the studies presented in **Chapters 3, 4 and 5**. Participants provided blood samples which were used for MS-based targeted protein quantification, as well as sociodemographic and other patient characteristics. Specific details of recruitment and selection criteria and diagnoses of psychiatric disorders can be found in the individual chapters. The research protocols of the studies were approved by the relevant ethical committees, written informed consent was obtained from all participants, and the studies were conducted under the standards of Good Clinical Practice and in compliance with the principles of the Declaration of Helsinki [298].

2.2. Blood samples

2.2.1. Serum

Serum samples were used to identify blood-based proteomic biomarkers in **Chapters 3 and 4**. Serum (and plasma) has been widely used as sample material of choice for for blood-based proteomic biomarker discovery. It contains approximately 60–80 mg/mL of proteins, as well as various small molecules including amino acids, lipids, salts, and sugars [219]. More than 10,000 proteins are estimated to be present in serum, in concentrations that span 10 orders of magnitude, resulting in one of the widest dynamic range for proteomes studied to date [217], [299]. However, a comprehensive characterisation of the human serum proteome is analytically challenging, as approximately 95% of the total protein content is accounted for by a small

number of high abundance proteins, such as albumin, immunoglobulins and complement factors [217], and the majority of the remaining proteins are present at relatively low concentrations. While proteins in the low and medium abundance range hold great potential as novel candidate biomarkers, their detection could be hindered by the masking and subsequent suppression by the high abundance proteins [300]. Therefore, in discovery experiments (non-targeted MS; global protein profiling), it may be necessary to introduce strategies that reduce the sample complexity (such as fractionation, enrichment and/or depletion) during the sample preparation workflow, and in doing so, improve the detection of the lower abundance proteins and enhance the coverage of the blood proteome [299], [301], [302]. However, if the analyte of interest is already known, these issues can be avoided by adopting a targeted approach such as MRM.

2.2.2. DBS

In **Chapter 5**, DBS samples were used to identify blood-based proteomic biomarkers. DBS presents a relatively novel and innovative sampling technique with several advantages over the traditional use of serum or plasma as the sample source of blood [303]–[305]. These include: easy and minimally invasive sample collection (*e.g.*, self-collection by means of a finger prick with no requirement for venipuncture); relatively small blood volume requirement; convenient sample storage and shipment (samples can be shipped using standard postage as analytes remain stable at room temperature over extended periods of time); and lower cost [202], [303], [306]. DBS are also exempt from dangerous goods regulations (biohazard) and represent a relatively low safety risk. In addition, DBS sampling provides an opportunity to investigate potential protein biomarkers in peripheral blood cells (*e.g.*, erythrocytes, monocytes, and thrombocytes), which are not present in serum (or plasma), representing a promising area of research.

Since DBS sampling was first introduced for newborn screening of phenylketonuria and other congenital metabolic disorders [307], the utility of this sampling technique has been further demonstrated in various fields, such as therapeutic drug monitoring [308] and HIV screening [309]. The implementation of DBS sampling in proteomic biomarker discovery and validation holds great promise for the development of a non-invasive and cost-effective diagnostic test for MDD and other psychiatric disorders, especially as the possibility of remote or home sample self-collection could facilitate patient recruitment which is notoriously challenging for

psychiatric disorders [202], [310]. This will allow DBS sampling to reach a larger target population without compromising sample integrity or encountering financial limitations, which often limit the conduct of large-scale clinical studies. However, despite these advantages of using DBS samples over traditional serum or plasma samples, it should be noted that the quality of DBS samples may be poor if the instructions regarding sampling, drying, and posting are not followed by the participants [306]. Furthermore, a recent study from the group demonstrated that targeted protein quantification in DBS can be achieved to a comparable extent to that in conventional serum [202]. Nevertheless, despite the increased attention received by DBS sampling over the past years and its clear potential for application in the field of psychiatry, this remains largely unexplored [310]. Additionally, as the characterisation of the DBS matrix is hindered by the same issues of sample complexity as in the case of serum, similar considerations should be made to improve the detection of the lower abundance proteins where necessary (as outlined above in **Section 2.2.1**).

2.3. Multiple reaction monitoring (MRM)

MRM (also known as selected reaction monitoring; SRM) is a targeted MS technique that enables highly sensitive and specific quantitation of pre-selected proteins. Its original application was in the measurement of small molecules such as metabolites or drugs [311]. In MRM, protein quantitation is based on the selected measurement of unique constituent peptides which have been selected as surrogates for and are produced from their corresponding proteins through proteolytic digestion [210]. MRM is most effectively used in a LC-MS system, in which the separation capabilities of LC are coupled with the analytical capabilities of MS [215]. Key advantages of MRM-MS are its ability to target specific peptides, its multiplexing capacity to monitor and analyse hundreds of peptides in a single LC-MS run, and true internal standardisation via the use of stable isotope-labelled (SIL) peptides (see **Section 2.3.3** for more details on internal standard) [210], [312]. The MRM-based approach using SIL peptides has been shown to be capable of highly multiplexed quantitation of proteins in plasma and serum samples [313], [314], as well as in DBS samples [315], [316].

MRM-MS was used in **Chapters 3, 4 and 5** to measure the abundances of pre-selected panels of proteins. The majority of the proteins was selected based on their previous associations with

psychiatric disorders as gathered from the scientific literature [294]. Target proteins were represented by unique peptide sequences (*i.e.*, surrogates). MRM-MS methods were reproduced and/or adapted from *Ozcan et al. (2017)* [202], a previously published study on MRM-based biomarker discovery from the group. This study provides details on MRM assay development, including targeted protein-peptide selection, transition selection and interference screening, and LC-MS/MS analysis. The researchers conducting the sample preparation and the LC-MS/MS analysis were blinded to the clinical status of the participants.

2.3.1. Sample preparation

An overview of the sample preparation process is shown in **Figure 2.1**. For both serum and DBS, samples were prepared in a 96-well plate format using a liquid handling robotic system (Biomex NX liquid handler; Beckman Coulter, High Wycombe, UK). Samples were diluted with ammonium bicarbonate (protein extraction), and dithiothreitol and iodoacetamide were used for disulphide bond reduction and cysteine alkylation, respectively. Proteins were digested overnight using trypsin, resulting in arginine or lysine at the C-terminus.

2.3.1.1. Serum

In **Chapters 3 and 4**, serum samples were stored at -80 °C prior to LC-MS/MS analysis. 5 µL serum samples in each well were diluted with 105 µL of 50 mM ammonium bicarbonate. This was followed by disulphide bond reduction using 20.5 µL of 32.5 mM dithiothreitol (final concentration 5 mM) and cysteine alkylation using 20.5 µL of 75 mM iodoacetamide (final concentration 10 mM). Samples were then digested overnight using trypsin (1:50 ratio of enzyme to protein weights).

2.3.1.2. DBS

In **Chapter 5**, DBS samples were stored at room temperature prior to LC-MS/MS analysis, each in a sealed storage bag containing a desiccant. 3 mm DBS discs were punched from the centre of a blood drop using a semi-automated puncher (pneumatic card punch; Analytical Sales and Services, Inc.; Flanders, New Jersey, USA) and transferred to a 96-well plate. Proteins were extracted from the samples using 40 µL of 50 mM ammonium bicarbonate. This was followed by disulphide bond reduction using 15 µL of 18.3 mM dithiothreitol (final concentration 5 mM) and cysteine alkylation using 15 µL of 46.6 mM iodoacetamide (final concentration 10 mM). Samples were then digested overnight using 15 µL trypsin (1:20 ratio

of enzyme to protein weights). Trypsin-digested peptides were enriched and purified using C-18 microfilter 96-well plates (FNSC18 from Glygen Corp., Columbia, MD, USA) on a Beckman Coulter Biomek NX workstation, and eluted from C-18 with 60% acetonitrile.

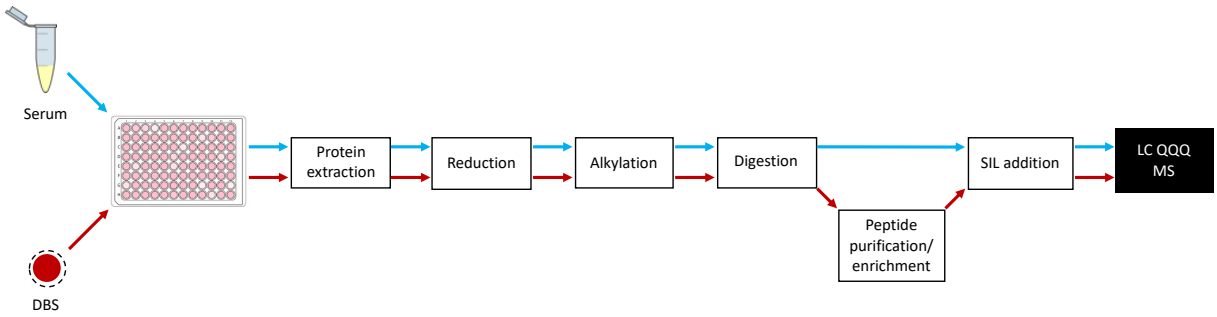


Figure 2.1. Workflow of automated sample preparation.
Adapted (with permission) from *Ozcan et al. (2017)*. Serum and DBS samples were prepared in a 96-well plate format using a liquid handling robotic system. Abbreviations: DBS (dried blood spot); LC QQQ-MS (liquid chromatography system coupled with triple quadrupole mass spectrometer); SIL (stable isotope-labelled peptide).

2.3.2. LC-MS/MS analysis

Triple quadrupole (QQQ) mass spectrometer is a tandem MS (MS/MS) instrument (a single instrument with two or more mass analysers) that is commonly used for MRM, in which peptide ions (precursors) undergo fragmentation to generate unique and characteristic fragment ions (products) [210]. The fragment ions containing the N- and C- terminal ends of the peptide ions are termed b- and y- ions, respectively. Essentially, the combinations of the intact peptide ions and their corresponding fragment ions comprise precursor-product ion pairs, also known as transitions, which are specific for the peptide sequences monitored [210].

A mass spectrometer typically consists of three principal components: an ionisation source to ionise the sample; mass analysers to separate ions based on their *m/z* ratios; and a detector to record the analyte signal [317]. Upon entering the mass spectrometer, peptides are introduced into the ionisation source where they become ionised. A QQQ mass spectrometer consists of two quadrupole mass filters in series with a collision cell in between. The first quardrupole (Q1) functions as a mass filter to selectively transmit precursor (peptide) ions of particular *m/z*

values (MS1) into the second quadrupole (Q2), where fragment ions are generated by collision-induced dissociation. The third quadrupole (Q3) functions as a second mass filter to selectively transmit fragment ions with high intensity of particular *m/z* values (MS2), and the detector measures the intensity of the precursor-product ion pairs (*i.e.*, transitions) [318]. MS/MS results are represented as mass spectra, in which the detected ion abundance (peak intensity) is plotted against the *m/z* value of the transitions. The monitoring of defined precursor-product ion pairs (transitions) enhances the signal-to-noise ratio and allows for a reliable determination of lower-abundance proteins [319]. A schematic of a QQQ mass spectrometer is shown in **Figure 2.2**.

In **Chapters 3, 4 and 5,** trypsin-digested peptides were separated and detected using a LC system coupled with a QQQ mass spectrometer (Agilent Infinity 1290 LC system and Agilent 6495 QQQ LC/MS system with Agilent Jet Stream electrospray ionisation (ESI) technology). Approximately 3.2 µg of digested serum or DBS peptides was injected into an Agilent AdvanceBio Peptide Map column (2.1 × 150 mm 2.7-micron) and separated at 50 °C. Peptides were eluted over a linear gradient from 3% to 30% acetonitrile in 0.1% formic acid in 45 minutes, at a flow rate of 0.3mL min^{-1}. The mass spectra were acquired in positive mode and the quantification of all analytes was carried out in MRM mode. Peptides were quantified at the transition level; in other words, the quantification of peptides was based on the signals of precursor-product ion pairs that reached the detector after the two-step mass filtering mechanism by the QQQ mass spectrometer. One to four interference-free transitions were selected for each peptide as described in *Ozcan et al. (2017)* [202]. Peptide retention times were identified using full scan data. Delta retention time window was 0.8 minutes and cycle time was 1 s.

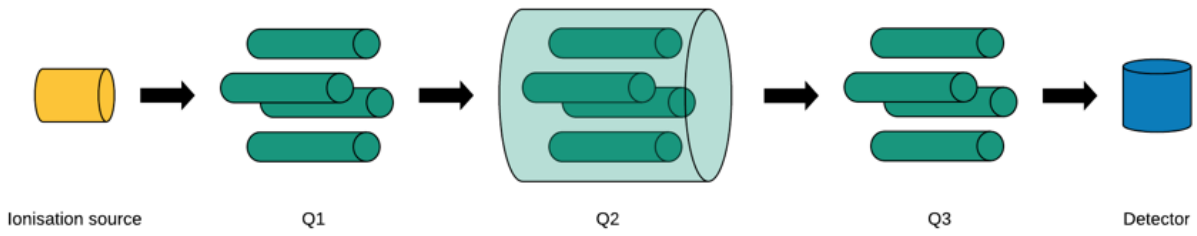


Figure 2.2. Schematic of QQQ mass spectrometer used for MRM.
Abbreviations: MRM (multiple reaction monitoring); QQQ (triple quadrupole).

2.3.3. Internal standard

SIL peptides were implemented as internal standards for each endogenous target peptide. This method, called stable isotope dilution, is the gold standard for peptide quantitation in MRM-MS [210], [320]. The SIL peptides are synthetically produced by incorporation of heavy (^{13}C- and ^{15}N- labelled) arginine or lysine. As they have the same sequence, and therefore the same chromatographic, ionisation and fragmentation properties as the endogenous peptides (but can be distinguished in the MS spectra), this technique facilitates the quantitation of endogenous peptides in the samples and improves the specificity and reproducibility of the measurements [202], [321]. SIL internal standard peptides were spiked into samples for each endogenous target peptide prior to LC-MS/MS analysis (that is, following trypsin digestion for serum samples and following peptide purification/enrichment for DBS samples; **Figure 2.1**). Endogenous and corresponding SIL peptide-transitions were monitored and acquired simultaneously by the QQQ mass spectrometer.

2.3.4. Quality control samples

In **Chapters 3, 4 and 5**, quality control (QC) samples were used to assess the technical variation associated with instrument performance and sample preparation. To assess the variation in instrument performance, a pooled QC sample was prepared by pooling together the digested clinical serum or DBS samples, and injected once a day along with the clinical samples for the duration of the entire study. To assess the variation in sample preparation, commercial serum samples (Human Sera S7023, Sigma Aldrich) were used in **Chapters 3 and 4**, and DBS samples obtained from a healthy volunteer (volunteer DBS) were used in **Chapter 5**. These QC samples were prepared following the same protocol as the clinical samples and distributed across the experimental plates.

2.4. Data pre-processing

2.4.1. Raw MS data processing

In **Chapters 3, 4 and 5**, raw MS data were processed using Skyline software package (version 3.1.0) [322]. Peptide-transition peaks were examined manually and peak integration boundaries were adjusted accordingly where necessary. Peak area values of the endogenous

and the SIL peptide-transitions were exported as a comma delimited data file for data pre-processing and analysis.

All statistical data pre-processing and analysis were carried out using R statistical software (version 3.6.0) [323], and all plots were made using the R package ggplot2 [324].

2.4.2. Quantifier transition selection

In **Chapters 3, 4, and 5**, quantifier transitions were selected for each peptide based on the abundance (peak area) values of the endogenous and the SIL peptide-transitions. The quantifier transition was selected as the transition that was consistently most abundant across the majority of the samples. Subsequent analyses were conducted using data of the quantifier transitions only. This approach has a number of advantages, among which the most important is that as the quantifier transition tends to be the most robustly measured transition for a given peptide, it would provide a robust prediction of disease outcome if selected as a biomarker/predictor [202]. The robustness of the most abundant peptide-transition measurements was previously demonstrated in *Ozcan et al. (2017)* [202].

2.4.3. Ratio normalisation

In **Chapters 3, 4 and 5**, peptide-transition quantification was based on the abundance ratio, calculated as the relative abundance of the endogenous and the SIL peptide-transitions. This process, referred to as ratio normalisation, is intended to account for any technical variation occuring across MS runs.

2.4.4. Logarithmic transformation

In **Chapters 3, 4 and 5**, the normalised peak area ratio vaules were $\log_2$-transformed for variance stabilisation prior to further analysis. While the variance of biological measurements often increases with intensity, logarithmic transformation decreases the variability of the data and improves normality, such that skewed distributions become more symmetric and the influence of high-abundance peptide-transitions is reduced [325], [326].

2.4.5. Principal component analysis

PCA is a technique used to reduce the dimensionality of a dataset and increase interpretability whilst retaining as much variation as possible [327]. It geometrically projects the data onto a

lower-dimensional space by identifying new variables, the principal components (PCs), which are linear combinations of the original variables and geometrically orthogonal (uncorrelated to each other) [328], [329]. The first PC is found as the direction along which the data shows the largest variation (in the space of the original variables); the second PC is found as the direction uncorrelated to the first PC along which the data shows the largest variation; subsequent PCs are found to successively maximise the variance of the data with the requirement of being uncorrelated to all previous PCs. The variation captured by the PCs decreases from the first PC to the last PC.

PCA was used in **Chapters 3, 4 and 5** to facilitate the visualisation of patterns or artefacts in the MS data and the identification of any outliers. It was conducted using the R package stats [323].

2.4.6. Coefficient of variation

In **Chapters 3, 4 and 5**, the geometric coefficient of variation (CV) was used to estimate the magnitude of variation in the MS data (in the original scale of measurement). For natural log-transformed data, the geometric CV was calculated as:

$$\mathrm{CV} = \sqrt{e^{sd^2-1}} \times 100\%$$

where *sd* = standard deviation of the log-transformed data.

The technical variation was assessed by calculating the geometric CVs of peptide-transition abundance ratios of the pooled QC samples (variation in MS instrument performance) and those of either the Sigma serum or volunteer DBS QC samples (variation in sample preparation) across MS runs. In addition, the biological variation was assessed by calculating the geometric CVs of peptide-transition abundance ratios of the clinical samples within each clinical sample group.

2.5. Statistical analysis

2.5.1. Model selection and bias-variance trade-off

Model selection is the process of selecting a statistical model from a set of potential models that best approximates the given data. In practice, this means identifying a set of model parameters that best constructs the relationship between the outcome (dependent variable) and the predictors (independent variables).

An important concept in model selection is the bias-variance trade-off [330]. Here, bias is the difference between the prediction of the model and the actual value, and variance is the variability of the model prediction for a given data point. Usually, simple models with few predictors have high bias and low variance, whereas complex models with many predictors have low bias and high variance (the statistical principle of parsimony) [290], [330], [331]. Underfitting occurs when there are not enough predictors in a model and the model is unable to capture the underlying pattern in the data. On the other hand, overfitting occurs when there are too many predictors in a model and the model captures not only the underlying pattern but also noise in the data. As a result, it has good predictive ability on the original data but this is not reproduced when applied to new data. Therefore, statistical model selection must seek an appropriate balance between overfitting and underfitting, which basically corresponds to the bias-variance trade-off.

2.5.2. LASSO regression

LASSO is a penalised regression method that simultaneously performs regularisation (shrinkage) and variable selection [281]. Logistic LASSO regression was used for binary classification problems in **Chapters 3, 4 and 5**.

2.5.2.1. Logistic regression

Logistic regression models the relationship between the probability of a binary outcome (*e.g.*, presence or absence of a disease) and explanatory variables (also called predictors or covariates). The probability (or likelihood) that the outcome is 1 ($Y = 1$) given that the explanatory variables are X can be written as:

$$\pi(X) = P(Y = 1|X)$$

The logistic (or sigmoid) function is [332]:

$$\pi(x_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})}}$$
$$i = 1, \ldots, n$$

where observations are indexed with subscript *i*; *n* is the number of observations; β_0 is the intercept (the bias term); and β_p are the regression coefficients for *p* explanatory variables. This function transforms the predictions into a binary outcome, which is expressed as a probability between 0 and 1. It can be manipulated to yield a model that is a linear combination of the explanatory variables:

$$\log\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$
$$\log\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$$
$$i = 1, \ldots, n$$

On the left-hand side is the logit or the logarithm of the odds (log odds), where the odds are the ratio of the probability that the event occurs versus the probability that the event does not occur. It is important to note that, while a covariate can be either a predictor variable, which is of direct interest, or an unwanted, confounding variable, since logistic regression examines the relationship between multiple covariates and the outcome, the resulting odds ratio is controlled for potential confounders. Hence, this is known as the adjusted odds ratio, as it has been adjusted for the other covariates, including potential confounders [266].

Maximum likelihood estimation is used to estimate the β parameters (regression coefficients) of a model. This involves identifying a set of parameters that maximises the probability (or likelihood) of observing the outcomes in the training dataset [332], [333]. The log-likelihood function, $\ell(\beta)$, is maximised (which is equivalent to minimising the negative log-likelihood function):

$$\ell(\beta) = \sum_{i=1}^{n} (y_i \log(\pi(x_i)) + (1 - y_i) \log(1 - \pi(x_i)))$$

The regression coefficients are interpreted as the estimated change in the log odds of the outcome associated with a one-unit change in the value of the independent variable, adjusting for all other independent variables in the model.

Several assumptions must be met when performing logistic regression [332]:

- Logistic regression does not assume that there is a linear relationship between the outcome variable and the explanatory variables (which is a key assumption of linear regression); however, it does assume that there is a linear relationship between the logit of the outcome and the explanatory variables.
- Logistic regression assumes that there is little or no multicollinearity among the explanatory variables, which means that they should not be highly correlated with each other.
- Logistic regression assumes that the observations are independent of each other.
- Logistic regression assumes that there are no strong influential outliers.
- Logistic regression requires that there is an adequate number of events per explanatory variable to avoid model overfitting; a minimum EPV of ten is commonly recommended as a rule-of-thumb (see **Section 1.4.2.2**).

2.5.2.2. Penalised logistic regression

When the number of observations (sample size) is small relative to the number of candidate predictors (small *n*, large *p*), standard regression methods may result in overfitted models which make inaccurate predictions [255]. This problem can be alleviated by using penalised regression methods, in which a penalty is introduced for complexity in the model. LASSO regularisation penalises the log-likelihood function (*i.e.*, adds a penalty term to the negative log-likelihood function) [281]:

$$\ell^*(\beta) = \ell(\beta) - \lambda \sum_{j=1}^{p} |\beta_j|$$

where λ is the regularisation (or shrinkage) parameter and p is the number of explanatory variables. By placing a constraint on the sum of the absolute values of the regression coefficients (L_1 regularisation), the coefficients of the variables that contribute poorly to the prediction are reduced exactly to zero, which effectively results in these variables being eliminated from the model. On the other hand, variables with non-zero coefficients are selected to be part of the model. Hence, the LASSO method simultaneously performs regularisation and variable selection to give sparse prediction models [281].

The extent of regularisation is controlled by the regularisation parameter λ. A larger λ value leads to more regularisation and results in a simpler model, as more variables are shrunken to zero (bias increases and variance decreases as λ increases). Note that when $\lambda = 0$, no regularisation is applied and the problem reduces to standard logistic regression. The optimal value of λ, and therefore the optimal model, is selected by K-fold cross-validation. More details on this are found in **Section 2.5.4** below.

By eliminating irrelevant variables from the model and retaining only the important variables, the LASSO method reduces model complexity and gives interpretable models. This also helps to reduce model overfitting and improve prediction accuracy, which is especially useful when the number of observations is small relative to the number of predictors. Therefore, LASSO is widely applied for developing clinical prediction models using high-dimensional datasets [255], [334].

However, LASSO has some limitations [335]. If $p > n$, then it selects at most n variables before saturating. This could potentially be problematic in cases where the number of important variables exceeds n, and due to the nature of the LASSO method, some important variables may fail to be selected in the model. Additionally, if there is a set of highly correlated variables in the data, then it tends to select only one variable from the set at random, which may limit the reproducibility of the results.

The R package glmnet was used in **Chapter 3** to implement LASSO [336].

2.5.3. Group LASSO regression

Group LASSO is an extension of the LASSO method which performs variable selection on predefined groups of variables [337]. This was used in **Chapters 4 and 5** in the presence of categorical variables (factors with more than two levels), to allow for sets of dummy variables derived from categorical variables to be selected together. The R package gglasso [338] was used to implement group LASSO.

2.5.4. *K*-fold cross-validation

K-fold cross-validation is a popular resampling technique for model selection and assessment that reduces overfitting. It was used to select the optimal value of the regularisation parameter λ for LASSO regression and group LASSO regression in **Chapters 3, 4 and 5**.

In *K*-fold cross-validation, the dataset is randomly partitioned into *K* folds, and each fold is retained as the test set whilst the remaining $K - 1$ folds are used as the training set. For each round of cross-validation, a model is fitted on the training set and its predictive performance (*i.e.*, misclassification error) is evaluated on the test set consisting of the held out observations for each value of λ. An overall measure of model predictive performance (mean cross-validation error) can be found for each value of λ by computing the average performance over all folds (**Figure 2.3**).

The one-standard-error rule, as suggested by Breiman et al. [339] and Hastie et al. [330], was used to select the most parsimonious (*i.e.*, simplest) model whose accuracy wass comparable with the most accurate model. While λ_{min} (left vertical dotted line in **Figure 2.3**) is the value of λ that gives minimum mean cross-validation error, λ_{1se} (right vertical dotted line in **Figure 2.3**) is the largest value of λ that gives a model whose error lies within one standard error of the minimum. The latter was chosen as the optimal value of λ as it resulted in a more regularised, and hence, parsimonious model. The model that was defined by the optimal value of the regularisation parameter was subsequently fit to the entire dataset [284].

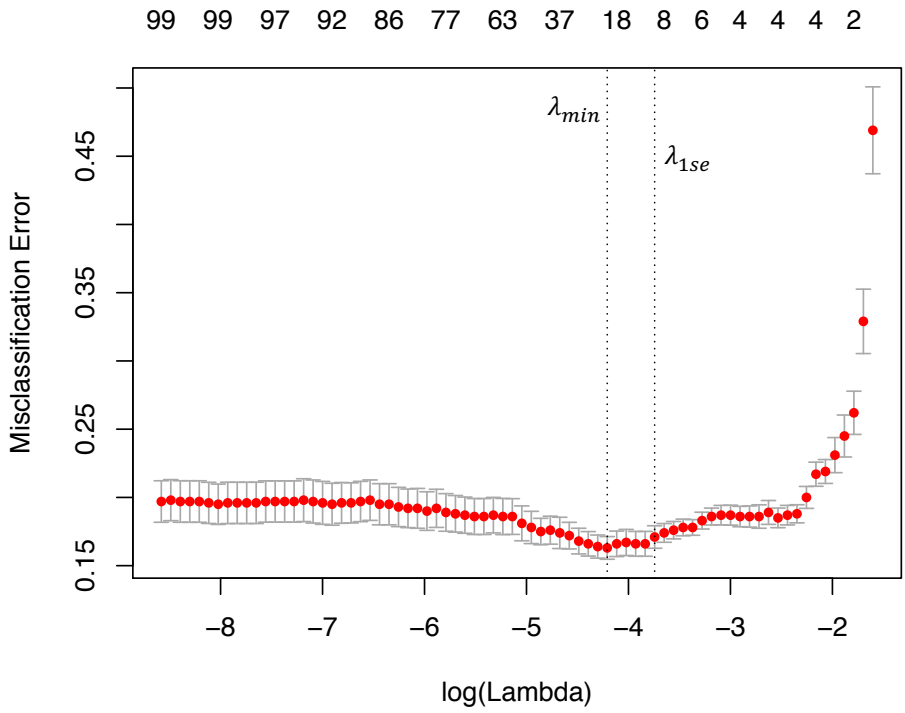


Figure 2.3. Selection of tuning parameter λ in LASSO using *K*-fold cross-validation.

Here, λ_{min} = 0.0149 gives mean cross-validation error = 0.163 ± 0.0083 [0.1547, 0.1713] and λ_{1se} = 0.0237 gives mean cross-validation error = 0.171 ± 0.0082 [0.1628, 0.1792].

2.5.5. Model predictive performance

In order to evaluate the predictive performance (or classification performance) of the models in **Chapters 3, 4 and 5**, receiver operating characteristic (ROC) curves were generated. The ROC curve plots the true positive rate (TPR), which measures the proportion of patients correctly identified by the model, against the false positive rate (FPR), which measures the proportion of controls incorrectly identified as patients by the model, at different classification thresholds. TPR (sensitivity) and FPR (1 – specificity) are calculated as:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

where TP, FP, TN and FN represent the number of true positives, false positives, true negatives and false negatives, respectively (**Table 2.1**). In other words, TP represents the number of patients correctly predicted as having the disease; FP represents the number of healthy individuals incorrectly predicted as having the disease (also known as type 1 error); TN represents the number of healthy individuals correctly predicted as being healthy; and FN represents the number of patients incorrectly predicted as being healthy (also known as type 2 error).

Table 2.1. Confusion matrix.

Abbreviations: FP (false positive); FN (false negative); TP (true positive); TN (true negative).

		Actual class	
		Patient	*Control*
Predicted class	*Patient*	TP	FP
	Control	FN	TN

The predictive performance was assessed using the area under the ROC curves (AUC). The AUC measures the likelihood that a randomly chosen individual with the disease (*i.e.*, patient) is assigned a higher predicted probability than a randomly chosen individual without the disease (*i.e.*, control) [340], [341]. A value of 1 reflects perfect classification performance, whereas a value of 0.5 reflects performance expected by random guessing and corresponds to a ROC curve along the $y = x$ line. In general, predictive performance is defined based on the following ranges of AUC values: 0.9 – 1 = excellent; 0.8 – 0.9 = good; 0.7 – 0.8 = fair; 0.6 – 0.7 = poor; 0.5 – 0.6 = fail [340], [341]. The R package ROCR was used to generate ROC curves and measure AUCs [342].

2.5.6. Univariate analysis

2.5.6.1. Mann-Whitney *U* test

The Mann-Whitney *U* test (also known as Wilcoxon's rank sum test) is a non-parametric test that is used to compare differences between two independent sample groups [343]. The

dependent variable is either continuous or ordinal (but not normally distributed), and the independent variable consists of two categorical groups. The null hypothesis is that there is no statistically significant difference between the two sample groups, and the alternative hypothesis is that there is a statistically significant difference between the sample groups.

2.5.6.2. Kruskal-Wallis *H* test

The Kruskal-Wallis *H* test is a non-parametric test that is used to compare differences between two or more independent sample groups [343]. The dependent variable is either continuous or ordinal (but not normally distributed), and the independent variable consists of two or more categorical groups. Similar to the Mann-Whitney *U* test, the null hypothesis is that there is no statistically significant difference between the sample groups, and the alternative hypothesis is that there is a statistically significant difference between the sample groups.

2.5.6.3. χ^2 test

The χ^2 test is a non-parametric test that is used to assess relationships between two categorical variables [343]. Each variable needs to have two or more categories, and the frequency of each category for one variable is compared across the categories of the second variable. The null hypothesis is that there is no statistically significant relationship between the categorical variables (*i.e.*, the categorical variables are independent), and the alternative hypothesis is that there is a statistically significant relationship between the categorical variables.

Univariate statistical tests were used in **Chapters 3, 4 and 5** to compare the sociodemographic and/or lifestyle characteristics of participants in different clinical groups. Mann-Whitney U tests were used to compare numerical variables, such as age and BMI, between two clinical groups; Kruskal-Wallis *H* tests were used to compare numerical variables between more than two clinical groups; and χ^2 tests were used to compare categorical variables, such as sex and education level, between two or more clinical groups. Additionally, in **Chapter 5**, Kruskal-Wallis *H* tests were used to assess for plate effects and Mann-Whitney *U* tests were used to assess for the effects of antidepressant medication use on participants' proteomic profiles and BMI. *P*-values < 0.05 were considered to be statistically significant. Univariate statistical tests were conducted using the R package stats [323].

Chapter 3 A multimodel approach for reproducible prediction model development

3.1. Introduction

Over the past few decades, considerable progress has been made in the field of biomarker research for psychiatric disorders, with many significant reports of potential candidate biomarker findings. Biomarkers have the potential to improve clinical practice, for instance, in relation to disease diagnosis, prognosis and treatment, especially given the symptom-based checklist approach to diagnosis and the 'trial-and-error' approach to treatment selection (see **Chapter 1**). Nonetheless, very few biomarkers have been successfully translated into FDA-approved clinical tests and are routinely applied in clinical settings [261], [262]. This is largely due to the lack of reproducibility of the research findings, which arises from the inherent complexity and heterogeneity of the disorders, as well as limitations in the pre-analytical, analytical and statistical workflow [15] (see **Chapter 1**).

While –omics technologies, such as LC-MS, enable hundreds or thousands of features (p) to be quantified, biomarker studies tend to involve a relatively small number of samples (n). This is especially the case for psychiatric disorders for which the recruitment of participants is notoriously difficult [193], [228]. As a result, prediction model development is typically performed on high-dimensional datasets (small n, large p). This can be problematic and lead to model overfitting, which occurs when the prediction model captures not only the underlying relationship of interest but also noise in the data [255]. Consequently, the predictive performance of the model is not reproduced when applied to new patient data; in other words, the model provides an overly optimistic assessment of the predictive performance based on the original data analysed [269]. More details on model overfitting can be found in **Chapter 1** (see **Section 1.4.2.2**).

Another important, yet often overlooked, issue that can limit the reproducibility of a prediction model is that of model selection uncertainty, which is when no single model is strongly

supported by the data. Model selection is traditionally considered to be a process of selecting a single model from a set of all possible candidate models that best approximates the data [283], [284]. Once the 'best' model has been selected, any uncertainty in model selection, which may arise from small changes in the data, is ignored. However, failure to account for this uncertainty may lead to biased inferences and predictions that are not reproducible, limiting the generalisability of the model [286], [287]. More details on model selection uncertainty can be found in **Chapter 1** (see **Section 1.4.2.3**).

Prediction models can potentially aid healthcare providers to make more statistically informed and objective decisions on the diagnosis, prognosis, and treatment of patients. However, as outlined above and in **Chapter 1**, the ability of the models to make accurate and reproducible predictions can be limited by problems such as model overfitting and model selection uncertainty. The aim of the work presented in this chapter is to explore the utility of a multimodel-based approach to predict the probability of having a complex psychiatric disorder while allowing for any uncertainty in model selection. The sensitivity of model selection to small changes in the data was evaluated by repeatedly applying LASSO regression with ten-fold cross-validation on the training set to obtain a set of 100 models. In the absence of a strongly supported model, that is, in the presence of model selection uncertainty, feature extraction and model averaging were applied across the 100 models to form weighted average prediction models [294]. Essentially, this approach enabled inferences and predictions to be derived from an entire set of models (*i.e.*, multimodel inference) instead of a single 'best' model [284], [289], [290], which improved model generalisability by not only accounting for model selection uncertainty but also reducing model overfitting [294]. Subsequently, the weighted average prediction models were applied to an independent test set to validate their predictive performance.

Although the present thesis is focused on depression, in this chapter, a serum MS dataset (77 proteins represented by 147 peptides) of 60 first-onset drug-naïve schizophrenia patients and 77 healthy controls was used to demonstrate proof-of-concept. This provided an ideal setting to investigate the utility of the multimodel approach for reproducible diagnostic prediction model development, not only as the patients were first-onset and drug-naïve, and schizophrenia tends to be more homogeneous than MDD as a disease in general, but also because using healthy controls as a reference group allowed for a clear-cut comparison, making the task of differentiation easier.

In the work reported in this chapter, blood samples were prepared by Nitin Rustogi, and run on the MS by Dr. Sureyya Ozcan. Raw MS data were processed by Dr. Sureyya Ozcan using Skyline. I was provided with the peak area values of the endogenous and SIL peptide-transitions and conducted all of the data pre-processing and analysis myself using R.

3.2. Materials and Methods

3.2.1. Clinical samples

The present work investigated participants from the Cologne Study, who were recruited by the Department of Psychiatry, University of Cologne, as described previously [344]. Blood serum samples were collected through venipuncture from 60 first-onset drug-naïve schizoprenia patients and 77 demographically-matched (for age and sex) healthy controls. Diagnosis of schizophrenia was based on the DSM-IV [345]. Participants with CNS disorders or other comorbidities were excluded. The ethical committees of the Medical Faculty of the University of Cologne and Addenbrooke's Hospital (Cambridge, UK) approved the protocols of this study, including procedures for sample collection and analysis. Informed consent was given in writing by all participants and the study was conducted according to the principles of the Declaration of Helsinki [298].

3.2.2. Targeted protein quantification

3.2.2.1. Sample preparation

Serum samples of the participants were randomised to allocate equal numbers of patients and controls across the experimental plates for LC-MS/MS analysis. Samples were prepared in a 96-well plate format as described in **Section 2.3.1.1**.

3.2.2.2. LC-MS/MS analysis

One hundred and forty-seven peptides representing 77 proteins, the majority previously associated with psychiatric disorders, were measured using targeted MRM-MS analysis. Peptides were quantified at the transition level. More details on LC-MS/MS analysis can be found in **Section 2.3.2.** The researchers conducting the sample preparation and the MS analysis were blinded to the clinical status of the participants.

3.2.2.3. QC samples

Pooled and Sigma serum QC samples were used to assess the technical variation associated with instrument performance and sample preparation as described in **Section 2.3.3**. A pooled QC sample, which was prepared by pooling together the digested clinical serum samples, was injected once a day along with the clinical samples for the duration of the entire study (11 injections in total). Eleven Sigma serum QC samples were prepared following the same protocol as the clinical samples and distributed across the plates.

3.2.3. MS data pre-processing

Pre-processing of the MS data was carried out as described in **Chapter 2** (see **Section 2.4**). Quantifier transitions were selected for each peptide based on the abundance values of the endogenous and the SIL peptide-transitions. A quantifier transition was selected as the transition that was consistently most abundant in at least 80% of the sample runs in both the endogenous and SIL peptides. Peptides for which quantifier transitions could not be selected based on these criteria, which can occur in the presence of competing transitions, the raw peak data were examined for manual selection. Subsequent analyses were conducted using data of the quantifier transitions only. Peptide quantification was based on the relative abundance of the endogenous and SIL peptide quantifier transitions, and the abundance ratio was $\log_2$-transformed for statistical analysis. There was no missing data.

3.2.4. Statistical quality control

PCA was conducted to identify any outliers based on the $\log_2$-transformed abundance ratios of 147 peptides (**Figure 3.1**). No outlier samples were identified.

In addition, the geometric CV was used to estimate the magnitude of variation in the proteomic data as described in **Chapter 2** (see **Section 2.4.6**). The median CV values for the pooled and Sigma serum QC runs, which were used to assess the technical variation, were 9.9% and 30.2%, respectively. The median CV values of the clinical samples, which were used to assess the biological variation, were 54.2% and 61.3% for the control and patient groups, respectively.

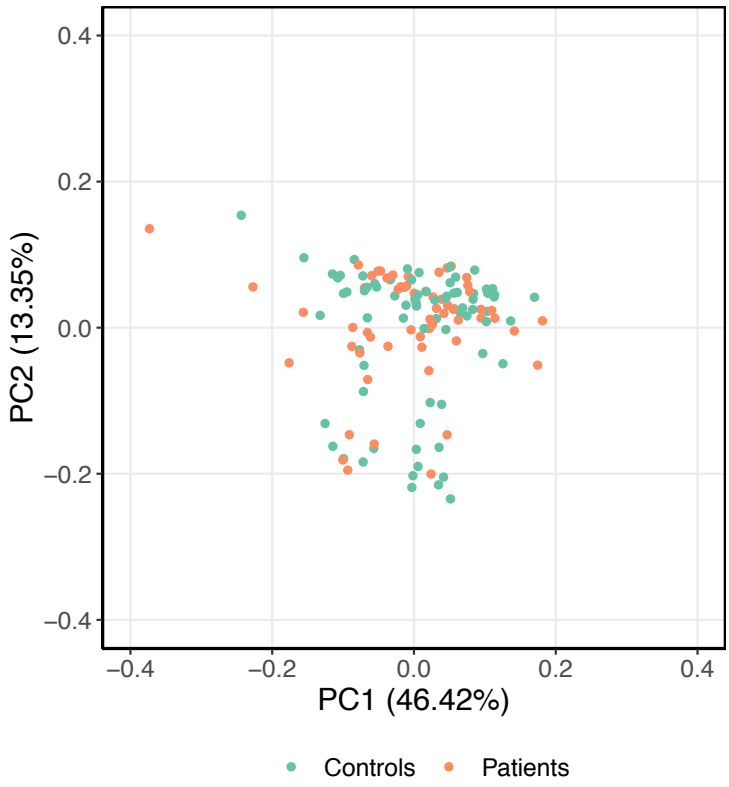


Figure 3.1. PCA plot of protein peptide abundance ratios of clinical samples.

The scores for the first two PCs are plotted with the percentage of variation accounted for by each PC shown in the axis labels. Data points are coloured by the clinical groups of the participants (60 first-onset drug-naïve schizophrenia patients and 77 healthy controls). Abbreviations: PC (principal component); PCA (principal component analysis).

3.2.5. Data analysis

The full dataset (60 first-onset drug-naïve schizophrenia patients and 77 controls) was randomly partitioned into 80:20 to form training (48 schizophrenia patients and 62 controls) and test (12 schizophrenia patients and 15 controls) sets. The training and test sets were matched for sex and age distribution. Sex and age of patients and controls were compared using a χ^2 test and a Mann-Whitney *U* test, respectively (see **Section 2.5.6**). *P*-values < 0.05 were considered to be statistically significant (**Table 3.1**).

3.2.5.1. Model selection

One hundred and forty-seven peptides representing 77 serum proteins were analysed for model selection to predict the probability of schizophrenia outcome in the training set. The full list of

the analysed proteins and peptides can be found in the **Appendix** (**Table A. 1**). LASSO regression with ten-fold cross-validation was used to reduce overfitting, such that the value of the shrinkage parameter λ that resulted in the most regularised model was selected. More details on LASSO regression and *K*-fold cross-validation can be found in **Chapter 2** (see **Sections 2.5.2 and 2.5.4)**.

To investigate model selection uncertainty, LASSO regression with ten-fold cross-validation was repeatedly applied on the training set 100 times to generate a set of 100 models. This allowed for an evaluation of the sensitivity of model selection to small changes in the data, which resulted from the random partitioning of the data in each application of ten-fold cross-validation. Models which comprised of the same combinations of features selected were grouped together.

3.2.5.2. Akaike information criterion

The model averaging framework was implemented using the Akaike information criterion (AIC), as described in Burnham and Anderson [283], [284]. The AIC measures how well a model approximates the data relative to the other possible models, with the best approximating model being the one with the lowest AIC value [346]. For a given model, the AIC can be expressed as a function of its maximised log-likelihood (L), which is a measure of model fit (see **Section 2.5.2**), and the number of estimated model parameters (k, *i.e.* the number of features selected in the model) [284], [346]:

$$\mathrm{AIC} = -2L + 2k$$

Therefore, the AIC measure gives a trade-off between model accuracy and model complexity, in other words, it seeks a balance between overfitting and underfitting. Essentially, as models with more estimated parameters (higher k) have higher AIC values, all other things being equal, models with fewer parameters are favoured [289].

In this work, given the small sample size (n) compared to the largest value of k ($n/k \leq 40$), the bias-corrected version of the AIC ($\mathrm{AIC_c}$) was adopted [347], [348]:

$$\mathrm{AIC_c} = \mathrm{AIC} + \frac{2k(k+1)}{n-k-1}$$

where n = sample size and k = number of estimated model parameters. In practice, since AIC_c converges to AIC as n increase, it is often recommended that AIC_c is used as default [347], [348]. The AIC_c was calculated for each model that was selected by LASSO regression with ten-fold cross-validation.

3.2.5.3. Akaike model weights

After model selection, the Akaike weight (w_i) was calculated for each model using the difference between its AIC_c and the lowest AIC (Δ_i). This used to assess the relative strength of the model among the candidate set of models [283], [284]:

$$\Delta_i = \mathrm{AIC_{c\,}}_i - \mathrm{AIC_{c\,min}}$$

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{r=1}^{R}\exp\left(-\frac{1}{2}\Delta_r\right)}$$

where w_i and $AIC_{c\,i}$ are, respectively, the Akaike weight and the AIC_c for model i in a set of R models, and $AIC_{c\,min}$ is the lowest AIC value across all models. The Akaike weight can be interpreted as the probability or the 'weight of evidence' in favour of a particular model being the best model approximating the outcome of interest given the set of models considered (*i.e.*, model probability) [283], [289]. It was a value between 0 and 1, and the sum of weights of all models was equal to 1:

$$\sum_{i=1}^{R} w_i = 1$$

3.2.5.4. Relative feature importance

Traditionally, assessment of the relative importance of features is often based on the best model alone, such as features selected in the best model are considered 'important', whereas those excluded from the best model are considered not important [283]. However, as this approach is too simplistic, relative feature importance can be refined by making inferences from all the models in the candidate set [283], [284]. In the present work, the relative importance of

individual features as well as that of subsets of features occurring together were assessed in two related but different ways:

i) Selection fraction

For each feature, the proportion of models out of 100 in which it was selected was measured. This was called the selection fraction, and was a value between 0 and 1. The relative importance of subsets of features occurring together was quantified by grouping together models which comprised of the same combinations of features selected and measuring the frequency of occurrence of each model group.

ii) Selection probability

Akaike model weights were used to estimate the probability that a particular feature was a component of the best model [284], called the selection probability (also more generally known as the predictor weight). The selection probability of each feature was calculated by summing the Akaike weights across the set of models in which it was selected, and was a value between 0 and 1. For example, the selection probability of a feature appearing in all of the high-ranking models would tend towards 1, whereas the selection probability of a feature only appearing in the very unlikely models would tend towards 0 [289]. In order to quantify the relative importance of subsets of features occurring together [284], for each model group, the Akaike weights of all corresponding models were summed to estimate the probability that the selected combination of features comprised the best approximating model [294].

3.2.5.5. Model selection uncertainty

The set of models obtained from the repeated application of LASSO regression with ten-fold cross-validation on the training set was evaluated to determine whether there existed uncertainty in model selection. If one of the model groups had $w > 0.9$, which could be interpreted as meaning that there was a 90% chance that it was the best model approximating the outcome given the candidate set of models, then this model could be considered as evidently superior to the other models [284]. In this instance, inference and prediction could be based on this model alone as there was minimal uncertainty in model selection, and accordingly, the coefficients of the features in this superior model group were estimated by averaging over the corresponding set of models.

However, when this was not the case (*i.e.*, there was no such model with $w > 0.9$) and there were other competing models that were equally or nearly as well supported as the best model, it would be disingenuous to only present the 'best' model as uncertainty in model selection was evident [289]. Under these circumstances, inference would need to be derived from the entire set of models to result in more reproducible predictions. Therefore, in order to obtain more reproducible predictions of the probability of having schizophrenia in the presence of model selection uncertainty, a multimodel approach was adopted by applying feature extraction and model averaging across the full set of 100 models to form weighted average prediction models.

3.2.5.6. Feature extraction

Feature extraction was used to define prediction models based on the relative importance of the features. Selection fraction and probability thresholds of 0.9 and 0.8 were used, such that only features with selection fractions or probabilities greater than or equal to the defined threshold were included in the models. Four prediction models were defined: a model consisting of features with selection fractions ≥ 0.9; a model consisting of features with selection fractions ≥ 0.8; a model consisting of features with selection probabilities ≥ 0.9; and a model consisting of features with selection probabilities ≥ 0.8. This feature extraction strategy provided the advantage of further reducing overfitting by excluding features that were less frequently selected or those with low selection probabilities based on Akaike model weights, which also had the effect of enhancing model interpretability.

3.2.5.7. Model averaging

Model averaging was used to estimate the weighted mean coefficient ($\widehat{\bar{\beta}}$) of a given feature of interest across the entire set of models:

$$\widehat{\bar{\beta}} = \sum_{i=1}^{R} w_i \hat{\beta}_i$$

where w_i and $\hat{\beta}_i$ are, respectively, the Akaike weight and the estimated feature coefficient in model i across a set of R models [284]. As all models were considered under this framework, those in which the feature of interest was not selected contributed nothing to the weighted mean coefficient estimate, resulting in the shrinkage of the coefficient towards zero [289]. As a result, poorly weighted features contributed less to the predictions than strongly weighted features.

3.2.5.8. Predictive performance

Predictive performance of the models when applied to the training and test sets was evaluated by plotting ROC curves and measuring the AUC. More details on model predictive performance can be found in **Chapter 2** (see **Section 2.5.5)**.

3.3. Results

One hundred and forty-seven peptides were measured in a training set of 48 first-onset drug-naïve schizophrenia patients and 62 healthy controls, and a test set of 12 first-onset drug-naïve schizophrenia patients and 15 healthy controls (**Table 3.1**). One hundred prediction models were obtained from repeatedly applying LASSO regression with ten-fold cross-validation on the training set. Akaike weights were calculated for each model to estimate the probability of it being the best model approximating schizophrenia outcome in the training set.

3.3.1. Model selection and Akaike model weights

Results across the 100 prediction models revealed some uncertainty in model selection (**Table 3.2**). The number of peptide features selected in a model ranged between one and 49, with a median of 12 (**Figure 3.2a**). Eight features out of 147 were selected at least 80 times out of 100, among which six features were selected at least 90 times (**Table 3.2**); one feature (HPT_VTSIQDWVQK) was selected 100 times. Fifty-five features were selected at least once (among which 25 features were selected less than 10 times out of 100), and 92 features were never selected, which was consistent with limited overfitting (**Figure 3.2b**). Moreover, feature selection probabilities were calculated to estimate the probability that a particular feature was a component of the best model. Nine features had selection probabilities ≥ 0.8, among which seven features had selection probabilities ≥ 0.9, and one feature (HPT_VTSIQDWVQK) had a selection probability of 1 (the 92 features that were never selected had selection probabilities of zero). The selection fraction and selection probability of the features were significantly correlated (Pearson's product-moment correlation coefficient = 0.97; p-value < 0.01; **Figure 3.2c**), very much as expected.

On average, the candidate set of models showed a good predictive performance on the training set (median AUC = 0.85; mean ± 95% confidence intervals AUC = 0.86 ± 0.01; **Figure 3.2d**).

Thirty model groups were identified based on the combination of features selected together (**Table 3.3**). For each model group, the frequency of occurrence and the probability of it being the best model were measured to assess the relative importance of subsets of features occurring together. The model group with the highest frequency of occurrence (20 times out of 100) consisted of 12 features and had the highest model probability of 0.35, which could be interpreted as there being a 35% chance that this model group was the best model. Given this relatively low weight ($w < 0.9$), there was insufficient evidence in favour of a superior model approximating schizophrenia outcome in the training set; in other words, there existed model selection uncertainty. When this model group was applied to predict the probability of schizophrenia outcome in the training set, it demonstrated a good predictive performance (AUC = 0.85; **Figure 3.3a**). However, its performance dropped to fair when applied to the test set (AUC = 0.73; **Figure 3.3b**), which was indicative of model overfitting. The remaining model groups each occurred eight times or fewer, and had model probabilities of 0.13 or smaller. Therefore, the considerable variability in feature selection and the lack of a strongly supported model approximating the probability of having schizophrenia in the training set highlighted the need for a multimodel approach.

3.3.2. Feature extraction and model averaging

In the presence of model selection uncertainty, a multimodel approach was adopted to form weighted average prediction models. Four weighted average prediction models (average selection fraction 0.9, average selection fraction 0.8, average selection probability 0.9, and average selection probability 0.8) were defined by applying feature extraction using selection fraction and selection probability thresholds of 0.9 and 0.8, followed by model averaging across all 100 models to obtain weighted mean coefficient estimates of the selected features. The first average model consisted of six features with selection fractions ≥ 0.9; the second average model consisted of eight features with selection fractions ≥ 0.8; the third average model consisted of seven features with selection probabilities ≥ 0.9; and the fourth average model consisted of nine features with selection probabilities ≥ 0.8 (**Table 3.4**). The weighted average models were applied to predict the probability of schizophrenia outcome in the training and test sets, and their predictive performance was evaluated. Overall, the models demonstrated a good predictive performance when applied to the training set (AUC = 0.81 – 0.84; **Figure 3.3a**), and a fair predictive performance when applied to the test set (AUC = 0.74 – 0.77; **Figure 3.3b**). Importantly, the weighted average models achieved lower training AUCs and higher test

AUCs than the best model group, which was expected, as the weighted average models consisted of fewer features. This demonstrated the effectiveness of the multimodel approach in alleviating model overfitting. Among the average models, comparisons between their predictive performance on the training and test sets indicated that model overfitting was most evident in the fourth model, which consisted of the largest number of features and achieved the highest training AUC and the lowest test AUC (0.84 and 0.74, respectively). On the other hand, model overfitting was least evident (*i.e.*, most effectively alleviated) in the first model, which consisted of the smallest number of features. This model achieved the lowest training AUC and the highest test AUC (0.81 and 0.77, respectively), demonstrating the smallest drop in predictive performance.

Table 3.1. Demographic characteristics of healthy controls and first-onset drug-naïve schizophrenia patients.

Age is shown as the mean (standard deviation) and sex is shown as the percentage of males and females for participants in the full (n = 137), training (n = 110) and test (n = 27) datasets. P-values are from comparisons between patients and controls in the full dataset using a $\chi 2$ test for sex (categorical variable) and a Mann-Whitney U test for age (numerical variable).

	Healthy controls	**First-onset drug-naïve schizophrenia patients**	***p*-value**
n			
Full	77	60	
Training	62	48	
Test	15	12	
Sex % (male/female)			
Full	55.8/44.2	51.7/48.3	0.6264
Training	54.8/45.2	52.1/47.9	
Test set	60/40	50/50	
Age (years)			
Full	31.8 (8.7)	30.9 (10.5)	0.3350
Training	31.8 (8.7)	30.9 (11.2)	
Test	31.8 (9.2)	30.8 (7.3)	

Table 3.2. Summary of the selected features.

For each of the 55 features selected at least once, the selection fraction, selection probability, coefficient in the best model group, and weighted mean coefficient, are reported. The features are shown in descending order of selection fraction. They are represented in a protein_peptide format, and their abbreviations can be found in the **Appendix** (**Table A. 1**). Ninety-two features out of 147 were never selected (not shown).

Feature	Selection fraction	Selection probability	Best model group coefficient	Weighted mean coefficient
(Intercept)	1	1	-0.2757	-0.2756
HPT_VTSIQDWVQK	1	1	0.2659	0.254
IC1_TNLESILSYPK	0.99	0.9951	0.3363	0.3172
APOA2_SPELQAEAK	0.98	0.9927	-0.2618	-0.2487
APOC3_DALSSVQESQVAQQAR	0.97	0.99	-0.0802	-0.0843
APOC1_EFGNTLEDK	0.96	0.9878	-0.0623	-0.0601
APOA4_IDQNVEELK	0.94	0.9841	-0.1722	-0.1644
FETUA_HTLNQIDEVK	0.88	0.9569	-0.1352	-0.1272
ITIH4_GPDVLTATVSGK	0.8	0.8596	0.0989	0.092
A2MG_NEDSLVFVQTDK	0.77	0.8295	0.116	0.1098
IGHA1_DASGVTFTWTPSSGK	0.67	0.7077	-0.0443	-0.0501
APOH_EHSSLAFWK	0.62	0.6591	0.02	0.039
IGHA2_DASGATFTWTPSSGK	0.42	0.3107	0	-0.0081
AACT_EQLSLLDR	0.41	0.6818	0.0181	0.0151
FCN3_YGIDWASGR	0.36	0.178	0	0.0053
A2AP_FDPSLTQR	0.3	0.0598	0	0.0044
C1RL_GSEAINAPGDNPAK	0.27	0.0398	0	0.0029
A1BG_ATWSGAVLAGR	0.22	0.0168	0	-0.0006
PHLD_NQVVIAAGR	0.22	0.0168	0	-0.0006
KNG1_DFVQPPTK	0.21	0.0138	0	-0.0029
IGHG2_TTPPMLDSDGSFFLYSK	0.17	0.0043	0	-0.0002
KLKB1_LSMDGSPTR	0.16	0.004	0	-0.0001
CBPB2_DTGTYGFLLPER	0.15	0.0001	0	0
CLUS_IDSLLENDR	0.15	0.0001	0	0

HRG_ADLFYDVEALDLESPK	0.15	0.0001	0	0
PEDF_TVQAVLTVPK	0.15	0.0001	0	0
A2AP_DSFHLDEQFTVPVEMMQAR	0.14	0.012	0	0.0006
HBG1_MVTAVASALSSR	0.14	0.0043	0	0.0001
ANT3_FDTISEK	0.12	0	0	0
ANGT_ALQDQLVLVAAK	0.11	0.0357	0	-0.0006
TTHY_VLDAVR	0.11	0	0	0
AMBP_ETLLQDFR	0.09	0	0	0
APOL1_LNILNNNYK	0.09	0	0	0
APOL1_VTEPISAESGEQVER	0.09	0	0	0
ITIH1_GSLVQASEANLQAAQDFVR	0.09	0	0	0
ITIH4_ETLFSVMPGLK	0.09	0	0	0
THRB_SGIECQLWR	0.09	0	0	0
TRFE_EGYYGYTGAFR	0.09	0	0	0
AACT_EIGELYLPK	0.08	0.043	0	0.0005
APOC2_ESLSSYWESAK	0.06	0	0	0
APOF_SLPTEDCENEK	0.06	0	0	0
CO2_HAIILLTDGK	0.06	0	0	0
CO6_TLNICEVGTIR	0.06	0	0	0
GELS_SEDCFILDHGK	0.06	0	0	0
A1BG_SGLSTGWTQLSK	0.05	0	0	0
A2MG_AIGYLNTGYQR	0.05	0.009	0	0
CO3_VYAYYNLEESCTR	0.05	0	0	0
A1AT_LSITGTYDLK	0.03	0	0	0
APOD_VLNQELR	0.03	0	0	0
APOE_LEEQAQQIR	0.03	0	0	0
CAH1_ADGLAVIGVLMK	0.03	0	0	0
CD5L_EATLQDCPSGPWGK	0.03	0	0	0
HBA_MFLSFPTTK	0.03	0	0	0
SHBG_IALGGLLFPASNLR	0.03	0	0	0
CO4A_ITQVLHFTK	0.01	0	0	0
LUM_SLEDLQLTHNK	0.01	0	0	0

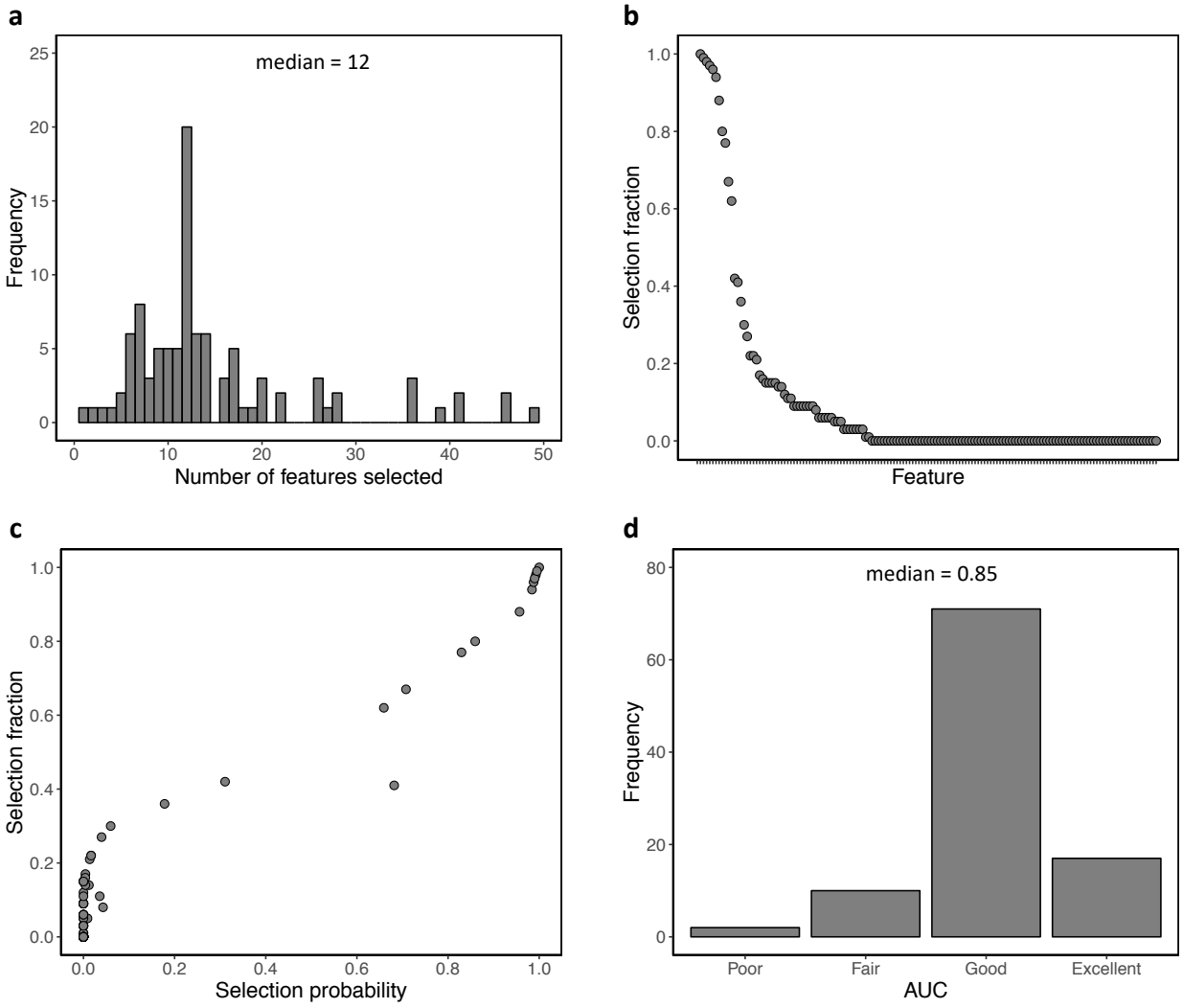


Figure 3.2. Summary of 100 models obtained from repeated application of LASSO regression with ten-fold cross-validation on the training set.

a) The number of features selected in each model. b) The selection fraction of each feature. c) The selection fraction and the selection probability of each feature. d) The AUC of each model. Abbreviations: AUC (area under the receiver operating characteristic curve); LASSO (least absolute shrinkage and selection operator).

Table 3.3. Model groups based on the combinations of features selected.

Thirty model groups were identified. For each model group, the frequency of occurrence and the model probability (summed Akaike weights) are reported. Model groups are shown in descending order of model probability. The features are represented in a protein_peptide format, and their abbreviations can be found in the **Appendix** (**Table A. 1**).

Model group	Number of features	Frequency	Model probability
A2MG_NEDSLVFVQTDK + AACT_EQLSLLDR + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + ITIH4_GPDVLTATVSGK	12	20	0.3484
A2MG_NEDSLVFVQTDK + AACT_EQLSLLDR + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + ITIH4_GPDVLTATVSGK	13	6	0.1327
A2MG_NEDSLVFVQTDK + AACT_EQLSLLDR + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + ITIH4_GPDVLTATVSGK	14	6	0.1183
A2MG_NEDSLVFVQTDK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR +	9	5	0.0990

FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + ITIH4_GPDVLTATVSGK			
APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK	7	8	0.0973
A2MG_NEDSLVFVQTDK + AACT_EQLSLLDR + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + ITIH4_GPDVLTATVSGK	11	5	0.0486
APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + ITIH4_GPDVLTATVSGK	8	3	0.0300
APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + HPT_VTSIQDWVQK + IC1_TNLESILSYPK	6	6	0.0272
A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + AACT_EIGELYLPK + ANGT_ALQDQLVLVAAK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + C1RL_GSEAINAPGDNPAK + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + ITIH4_GPDVLTATVSGK	17	5	0.023
A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + AACT_EIGELYLPK + AACT_EQLSLLDR + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK +	16	3	0.0200

APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + ITIH4_GPDVLTATVSGK			
A2MG_NEDSLVFVQTDK + AACT_EQLSLLDR + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + ITIH4_GPDVLTATVSGK	10	1	0.0139
A2MG_AIGYLNTGYQR + A2MG_NEDSLVFVQTDK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + ITIH4_GPDVLTATVSGK	10	4	0.009
A1BG_ATWSGAVLAGR + A2AP_DSFHLDEQFTVPVEMMQAR + A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + ANGT_ALQDQLVLVAAK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + C1RL_GSEAINAPGDNPAK + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + ITIH4_GPDVLTATVSGK + KNG1_DFVQPPTK + PHLD_NQVVIAAGR	20	3	0.0077
HPT_VTSIQDWVQK	1	1	0.0049
A1BG_ATWSGAVLAGR + A2AP_DSFHLDEQFTVPVEMMQAR + A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR +	22	1	0.0039

APOH_EHSSLAFWK + C1RL_GSEAINAPGDNPAK + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + HBG1_MVTAVASALSSR + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + IGHG2_TTPPMLDSDGSFFLYSK + ITIH4_GPDVLTATVSGK + KLKB1_LSMDGSPTR + KNG1_DFVQPPTK + PHLD_NQVVIAAGR			
APOA2_SPELQAEAK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + HPT_VTSIQDWVQK + IC1_TNLESILSYPK	5	2	0.0036
A1BG_ATWSGAVLAGR + A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + ANGT_ALQDQLVLVAAK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + C1RL_GSEAINAPGDNPAK + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + ITIH4_GPDVLTATVSGK + PHLD_NQVVIAAGR	18	1	0.0030
APOA2_SPELQAEAK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK	3	1	0.0027
HPT_VTSIQDWVQK + IC1_TNLESILSYPK	2	1	0.0024
APOA2_SPELQAEAK + APOC3_DALSSVQESQVAQQAR + HPT_VTSIQDWVQK + IC1_TNLESILSYPK	4	1	0.0022
A1BG_ATWSGAVLAGR + A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + ANGT_ALQDQLVLVAAK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + C1RL_GSEAINAPGDNPAK + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK +	19	1	0.0018

ITIH4_GPDVLTATVSGK + KNG1_DFVQPPTK + PHLD_NQVVIAAGR			
A1BG_ATWSGAVLAGR + A2AP_DSFHLDEQFTVPVEMMQAR + A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + ANGT_ALQDQLVLVAAK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + C1RL_GSEAINAPGDNPAK + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + HBG1_MVTAVASALSSR + HPT_VTSIQDWVQK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + IGHG2_TTPPMLDSDGSFFLYSK + ITIH4_GPDVLTATVSGK + KNG1_DFVQPPTK + PHLD_NQVVIAAGR	22	1	< 0.0001
A1BG_ATWSGAVLAGR + A2AP_DSFHLDEQFTVPVEMMQAR + A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + C1RL_GSEAINAPGDNPAK + CBPB2_DTGTYGFLLPER + CLUS_IDSLLENDR + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + HBG1_MVTAVASALSSR + HPT_VTSIQDWVQK + HRG_ADLFYDVEALDLESPK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + IGHG2_TTPPMLDSDGSFFLYSK + ITIH4_GPDVLTATVSGK + KLKB1_LSMDGSPTR + KNG1_DFVQPPTK + PEDF_TVQAVLTVPK + PHLD_NQVVIAAGR	26	3	0.0001
A1AT_LSITGTYDLK + A1BG_ATWSGAVLAGR + A2AP_DSFHLDEQFTVPVEMMQAR + A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + AMBP_ETLLQDFR + ANT3_FDTISEK + APOA2_SPELQAEAK + APOA4_IDQNVEELK +	36	3	< 0.0001

APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + APOL1_LNILNNNYK + APOL1_VTEPISAESGEQVER + C1RL_GSEAINAPGDNPAK + CBPB2_DTGTYGFLLPER + CLUS_IDSLLENDR + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + HBG1_MVTAVASALSSR + HPT_VTSIQDWVQK + HRG_ADLFYDVEALDLESPK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + IGHG2_TTPPMLDSDGSFFLYSK + ITIH1_GSLVQASEANLQAAQDFVR + ITIH4_ETLFSVMPGLK + ITIH4_GPDVLTATVSGK + KLKB1_LSMDGSPTR + KNG1_DFVQPPTK + PEDF_TVQAVLTVPK + PHLD_NQVVIAAGR + THRB_SGIECQLWR + TRFE_EGYYGYTGAFR + TTHY_VLDAVR			
A1BG_ATWSGAVLAGR + A1BG_SGLSTGWTQLSK + A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + AMBP_ETLLQDFR + ANT3_FDTISEK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC2_ESLSSYWESAK + APOC3_DALSSVQESQVAQQAR + APOD_VLNQELR + APOE_LEEQAQQIR + APOF_SLPTEDCENEK + APOH_EHSSLAFWK + APOL1_LNILNNNYK + APOL1_VTEPISAESGEQVER + C1RL_GSEAINAPGDNPAK + CAH1_ADGLAVIGVLMK + CBPB2_DTGTYGFLLPER + CD5L_EATLQDCPSGPWGK + CLUS_IDSLLENDR + CO2_HAIILLTDGK + CO3_VYAYYNLEESCTR + CO6_TLNICEVGTIR + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + GELS_SEDCFILDHGK + HBA_MFLSFPTTK + HPT_VTSIQDWVQK + HRG_ADLFYDVEALDLESPK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + IGHG2_TTPPMLDSDGSFFLYSK +	46	2	< 0.0001

ITIH1_GSLVQASEANLQAAQDFVR + ITIH4_ETLFSVMPGLK + ITIH4_GPDVLTATVSGK + KLKB1_LSMDGSPTR + KNG1_DFVQPPTK + PEDF_TVQAVLTVPK + PHLD_NQVVIAAGR + SHBG_IALGGLLFPASNLR + THRB_SGIECQLWR + TRFE_EGYYGYTGAFR + TTHY_VLDAVR			
A1BG_ATWSGAVLAGR + A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + AMBP_ETLLQDFR + ANT3_FDTISEK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC2_ESLSSYWESAK + APOC3_DALSSVQESQVAQQAR + APOF_SLPTEDCENEK + APOH_EHSSLAFWK + APOL1_LNILNNNYK + APOL1_VTEPISAESGEQVER + C1RL_GSEAINAPGDNPAK + CBPB2_DTGTYGFLLPER + CLUS_IDSLLENDR + CO2_HAIILLTDGK + CO6_TLNICEVGTIR + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + GELS_SEDCFILDHGK + HBG1_MVTAVASALSSR + HPT_VTSIQDWVQK + HRG_ADLFYDVEALDLESPK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + IGHG2_TTPPMLDSDGSFFLYSK + ITIH1_GSLVQASEANLQAAQDFVR + ITIH4_ETLFSVMPGLK + ITIH4_GPDVLTATVSGK + KLKB1_LSMDGSPTR + KNG1_DFVQPPTK + PEDF_TVQAVLTVPK + PHLD_NQVVIAAGR + THRB_SGIECQLWR + TRFE_EGYYGYTGAFR + TTHY_VLDAVR	39	1	< 0.0001
A1BG_ATWSGAVLAGR + A2AP_DSFHLDEQFTVPVEMMQAR + A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + ANT3_FDTISEK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + C1RL_GSEAINAPGDNPAK + CBPB2_DTGTYGFLLPER + CLUS_IDSLLENDR + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK +	28	2	< 0.0001

HBG1_MVTAVASALSSR + HPT_VTSIQDWVQK + HRG_ADLFYDVEALDLESPK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + IGHG2_TTPPMLDSDGSFFLYSK + ITIH4_GPDVLTATVSGK + KLKB1_LSMDGSPTR + KNG1_DFVQPPTK + PEDF_TVQAVLTVPK + PHLD_NQVVIAAGR + TTHY_VLDAVR			
A1BG_ATWSGAVLAGR + A1BG_SGLSTGWTQLSK + A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + AMBP_ETLLQDFR + ANT3_FDTISEK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC2_ESLSSYWESAK + APOC3_DALSSVQESQVAQQAR + APOF_SLPTEDCENEK + APOH_EHSSLAFWK + APOL1_LNILNNNYK + APOL1_VTEPISAESGEQVER + C1RL_GSEAINAPGDNPAK + CBPB2_DTGTYGFLLPER + CLUS_IDSLLENDR + CO2_HAIILLTDGK + CO3_VYAYYNLEESCTR + CO6_TLNICEVGTIR + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + GELS_SEDCFILDHGK + HBG1_MVTAVASALSSR + HPT_VTSIQDWVQK + HRG_ADLFYDVEALDLESPK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + IGHG2_TTPPMLDSDGSFFLYSK + ITIH1_GSLVQASEANLQAAQDFVR + ITIH4_ETLFSVMPGLK + ITIH4_GPDVLTATVSGK + KLKB1_LSMDGSPTR + KNG1_DFVQPPTK + PEDF_TVQAVLTVPK + PHLD_NQVVIAAGR + THRB_SGIECQLWR + TRFE_EGYYGYTGAFR + TTHY_VLDAVR	41	2	< 0.0001
A1BG_ATWSGAVLAGR + A1BG_SGLSTGWTQLSK + A2AP_FDPSLTQR + A2MG_AIGYLNTGYQR + A2MG_NEDSLVFVQTDK + AMBP_ETLLQDFR + ANT3_FDTISEK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC2_ESLSSYWESAK +	49	1	< 0.0001

APOC3_DALSSVQESQVAQQAR + APOD_VLNQELR + APOE_LEEQAQQIR + APOF_SLPTEDCENEK + APOH_EHSSLAFWK + APOL1_LNILNNNYK + APOL1_VTEPISAEGEQVER + C1RL_GSEAINAPGDNPAK + CAH1_ADGLAVIGVLMK + CBPB2_DTGTYGFLLPER + CD5L_EATLQDCPSGPWGK + CLUS_IDSLLENDR + CO2_HAIILLTDGK + CO3_VYAYYNLEESCTR + CO4A_ITQVLHFTK + CO6_TLNICEVGTIR + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + GELS_SEDCFILDHGK + HBA_MFLSFPTTK + HPT_VTSIQDWVQK + HRG_ADLFYDVEALDLESPK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + IGHG2_TTPPMLDSDGSFFLYSK + ITIH1_GSLVQASEANLQAAQDFVR + ITIH4_ETLFSVMPGLK + ITIH4_GPDVLTATVSGK + KLKB1_LSMDGSPTR + KNG1_DFVQPPTK + LUM_SLEDLQLTHNK + PEDF_TVQAVLTVPK + PHLD_NQVVIAAGR + SHBG_IALGGLLFPASNLR + THRB_SGIECQLWR + TRFE_EGYYGYTGAFR + TTHY_VLDAVR			
A1BG_ATWSGAVLAGR + A2AP_DSFHLDEQFTVPVEMMQAR + A2AP_FDPSLTQR + A2MG_NEDSLVFVQTDK + ANT3_FDTISEK + APOA2_SPELQAEAK + APOA4_IDQNVEELK + APOC1_EFGNTLEDK + APOC3_DALSSVQESQVAQQAR + APOH_EHSSLAFWK + C1RL_GSEAINAPGDNPAK + CBPB2_DTGTYGFLLPER + CLUS_IDSLLENDR + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + HBG1_MVTAVASALSSR + HPT_VTSIQDWVQK + HRG_ADLFYDVEALDLESPK + IC1_TNLESILSYPK + IGHA1_DASGVTFTWTPSSGK + IGHA2_DASGATFTWTPSSGK + IGHG2_TTPPMLDSDGSFFLYSK + ITIH4_GPDVLTATVSGK + KLKB1_LSMDGSPTR +	27	1	< 0.0001

KNG1_DFVQPPTK + PEDF_TVQAVLTVPK + PHLD_NQVVIAAGR

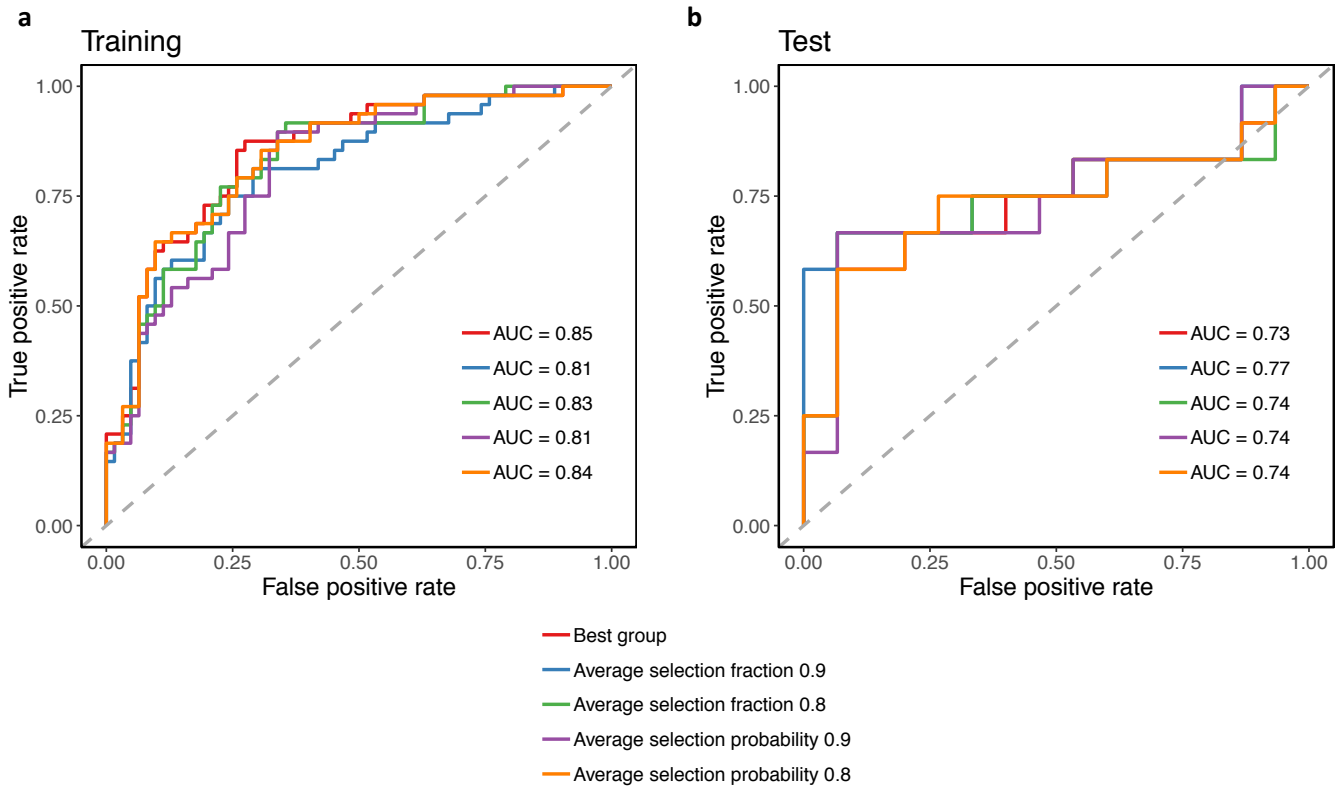


Figure 3.3. ROC curves showing model performance in predicting the probability of schizophrenia outcome.

The best group prediction model (*i.e.*, the model group with the highest frequency of occurrence and model probability) and four weighted average prediction models were applied to predict the probability of having schizophrenia in: a) the training set (48 first-onset drug-naïve schizophrenia patients *vs* 62 healthy controls), and b) the test set (12 first-onset drug-naïve schizophrenia patients *vs* 15 healthy controls). Abbreviations: AUC (area under the curve); ROC (receiver operating characteristic).

Table 3.4. Summary of the evaluated prediction models.

The best group prediction model was defined as the model group with the highest frequency of occurrence and model probability among the candidate set of models. Four weighted average prediction models (average selection fraction 0.9, average selection fraction 0.8, average selection probability 0.9, and average selection probability 0.8) were defined by applying feature extraction using selection fraction and selection probability thresholds of 0.9 and 0.8, followed by model averaging across all 100 models. For each of the evaluated models, the number of features selected, the predictive performance (as measured by the AUC) and the predictive accuracy when applied to the training and test sets, are shown. Abbreviations: AUC (area under the curve).

Model	Number of features	Training AUC	Training accuracy	Test AUC	Test accuracy
Best group	12	0.8498	0.7818	0.7333	0.7407
Average selection fraction 0.9	6	0.8071	0.7273	0.7722	0.8148
Average selection fraction 0.8	8	0.8310	0.7273	0.7444	0.7407
Average selection probability 0.9	7	0.8122	0.7091	0.7389	0.7037
Average selection probability 0.8	9	0.8424	0.7818	0.7389	0.6667

3.4. Discussion

This chapter explored and further developed the multimodel framework in the context of obtaining reproducible predictions of having a complex psychiatric disorder based a proteomic biomarker dataset. One hundred and forty-seven peptides representing 77 serum proteins were measured in 60 first-onset drug-naïve schizophrenia patients and 77 demographically-matched controls using targeted MRM-MS. Given the distinct clinical conditions of the participants under comparison (*i.e.*, first-onset drug-naïve patients *vs* healthy controls), the schizophrenia dataset represented an ideal scenario to establish proof-of-concept for investigating model selection uncertainty and demonstrating the advantage of the multimodel approach, despite the focus of the present thesis on depression.

To this end, the sensitivity of model selection to small changes in the data was evaluated by repeatedly applying LASSO regression with ten-fold cross-validation on the training set to obtain 100 models. Relative model weights (Akaike weights) were subsequently calculated for each model based on the AIC_c, such that the weights depended on how much the model was supported by the data given the candidate set of models [289]. Thus, by comparing multiple models in terms of approximating the data, the AIC_c enabled model selection uncertainty to be quantified and relative model weights to be determined in a relatively practical and simple manner [289], [290]. As no single model stood out as being the best model approximating schizophrenia outcome in the training set, a multimodel approach was adopted to produce inferences and predictions that were unconditional on a specific model, and instead derived across the entire set of models. Hence, model selection could be viewed as a way to obtain model weights for model averaging, rather than just a way to select only one model [283]. In addition, the fact that 92 out of 147 features were never selected in any of the 100 models indicates that the method used for model selection (*i.e.*, LASSO regression with ten-fold cross-validation) was effective in limiting model overfitting. It should be noted that model averaging using the AIC_c is an established method for deriving inferences and predictions from a set of models rather than a single 'best' model in the presence of model selection uncertainty [284], [289], [290].

A clear advantage of the multimodel approach, particularly given the considerable variability in feature selection across the 100 models, was that it enabled the risk of selecting one of the

less probable models by chance to be reduced [294]. Additionally, feature extraction to identify the most important features for the average model and exclude less important ones had the effect of further alleviating model overfitting and improving model interpretability. The relative importance of the features was assessed by measuring their selection fractions (the proportion of models out of 100 in which a particular feature was selected) and selection probabilities (the probability that a particular feature was a component of the best model, calculated using relative model weights). Weighted average prediction models, which were defined by applying feature extraction with selection fraction and selection probability thresholds of 0.9 and 0.8, followed by model averaging across all 100 models, consisted of between six and nine features, and demonstrated a good predictive performance when applied to the training set (AUC = 0.81 – 0.84), and a fair predictive performance when applied to the test set (AUC = 0.74 – 0.77) in differentiating between patients and controls. The most reproducible predictions of the probability of having schizophrenia were obtained by the simplest average model consisting of six features with selection fractions ≥ 0.9. Given the training and test set AUCs of 0.81 and 0.77, respectively, the extent of model overfitting appeared to be relatively small. In this model, six serum proteins, apolipoprotein A-II (APOA2), apolipoprotein A-IV (APOA4), apolipoprotein C-I (APOC1), apolipoprotein C-III (APOC3), haptoglobin (HPT) and plasma protease C1 inhibitor (IC1), were selected as predictors of schizophrenia outcome; among these, two proteins (HPT and IC1) were upregulated in schizophrenia patients compared to controls, whereas four proteins (APOA2, APOC3, APOC1, and APOA4) were downregulated in schizophrenia patients compared to controls. All of the selected proteins except for IC1 have been previously linked to schizophrenia, with the most robust finding for increased HPT levels in schizophrenia patients compared to controls [349]–[352].

Therefore, a parsimonious diagnostic prediction model, which produced the most reproducible predictions of schizophrenia outcome with as few features as possible, was obtained using the selection fraction threshold of 0.9 for feature extraction. Increasing model complexity by adding more features to the model had the effect of increasing overfitting and reducing generalisability (as the predictive performance on the training set was increased whereas that on the test set was reduced). Simple models tend to be favoured over complex ones as they are not only more interpretable but also potentially more cost-effective and convenient for application in a clinical setting. While both selection fractions and probabilities were used for feature extraction, the results indicate that the former may be a more favoured approach as it

led to simpler models in both cases of using thresholds of 0.9 and 0.8, although this finding may be specific to the dataset analysed. Note that in future applications of the multimodel approach, a lower threshold may be required for feature extraction if no features have selection fractions ≥ 0.9. Moreover, the advantage of the multimodel approach was further exemplified by the fact that presenting only the best model group when the amount of support for it being superior to the other models was insufficient ($w < 0.9$), and thereby ignoring model selection uncertainty, resulted in less reproducible predictions of schizophrenia outcome. This model group, which represented the best model that could be obtained from the set of 100 models, was more complex (*i.e.*, consisted of more features) and overfitted than the weighted average models. Therefore, these results suggest that it would be unlikely for predictions obtained from any of the 100 models without applying feature extraction and model averaging to outperform those obtained from the multimodel approach.

There are several limitations to the present study. While the predictive performance of the parsimonious model could be considered as reproducible in that there was a modest discrepancy between the training and test set AUCs, there is potential for improvement. More specifically, achieving a test AUC of at least 0.8 ('good' predictive performance, by definition [340], [341]) would be desirable considering the application of such a diagnostic prediction model in a clinical setting. Moreover, although the original dataset was partitioned to create training and test sets in order to allow for the prediction models to be externally validated on an independent dataset (*i.e.*, dataset other than that used for model training), the models could not be validated on an independent cohort of patients and controls recruited from a different clinical centre due to the lack of availability of such a dataset. Finally, it should be noted that since the clinical conditions of the investigated participants were very distinct in order to demonstrate proof-of-concept, the application of the presented multimodel approach to the development diagnostic prediction models using more clinically relevant datasets such as those consisting of MDD patients could potentially result in predictions of outcome that are not as good as those achieved in this work.

In conclusion, this chapter demonstrated the utility of a multimodel-based approach in addressing model overfitting and model selection uncertainty and thereby making reproducible predictions of the diagnosis of a complex psychiatric disorder using high-dimensional serum proteomic biomarker data. A parsimonious and generalisable diagnostic prediction model, comprised of the most important and fewest set of features and their weighted average

coefficient estimates, was developed in the presence of model selection uncertainty by the combined implementation of feature extraction and model averaging based on a set of candidate models.

Chapter 4 Integrating proteomic, sociodemographic and clinical data to predict future depression diagnosis in subthreshold symptomatic individuals

4.1. Introduction

MDD is a complex and burdensome disorder that is characterised by low mood and energy levels, as well as concentration problems, sleep disturbances and changes in weight and appetite [1], [353]. It affects more than 300 million people worldwide [2], and is estimated to become the most debilitating disorder worldwide by 2030 [11]. Diagnosis of MDD currently relies on the evaluation of symptoms in clinical interviews according to the criteria outlined in the DSM-5 [1] or ICD-11 [16] (see **Chapter 1**).

As discussed in **Chapter 1**, there is an increased interest in identifying individuals experiencing subthreshold depression [47], [55]–[61]. Diagnosis of MDD based on the DSM-5 requires individuals to present with at least five out of nine depressive symptoms, including at least one core symptom of depressed mood or anhedonia (loss of interest or pleasure), and the symptoms need to be present most of the day, nearly every day, for at least two weeks (**Table 1.1**) [1]. Under such a checklist diagnostic approach, those who present with depressive symptoms that do not fulfil the diagnostic criteria (*i.e.*, fewer than five symptoms and/or duration of symptoms for less than two weeks and/or lack of a core symptom) are overlooked, despite the converging evidence that the presence of subthreshold depressive symptoms is associated with increased functional and social impairment, reduced quality of life, and increased utilisation of health services [58], [60], [64], [67], [68]. Importantly, while subthreshold depression has been identified as a risk factor for developing MDD in the future [67], [72], [76], there is also growing evidence that targeting subthreshold individuals with indicated preventive

interventions can help to prevent or delay the onset of MDD [78]–[80]. This highlights the clinical importance of finding early manifestations or biomarkers of incident MDD in subthreshold individuals, which could be used to identify those who will benefit most from appropriate preventive interventions. Early and more accurate detection of MDD is also essential for reducing the disease burden and the related healthcare costs. More on subthreshold depression can be found in **Section 1.1.5**.

The probability of an individual having or developing MDD can be predicted using a risk prediction model, based on his or her molecular, sociodemographic and/or clinical characteristics [254], [255]. The clinical utility of such a model is in aiding the process of decision-making with regards to the diagnosis or treatment of patients or symptomatic help-seekers, and it is important that the performance of the model is reproduced when applied to new patient data (see **Chapter 1**). However, as the development of a prediction model in biomarker studies often involves performing model selection on high-dimensional data, model reproducibility can be limited by problems such as overfitting as well as model selection uncertainty. These problems, which are discussed in detail in **Chapter 3**, need to be appropriately addressed to ensure that a robust and generalisable model is obtained.

A standard approach employed by many biomarker studies is to use healthy controls as a reference population against which patients are compared. However, in this chapter, a more clinically relevant and appropriate sample population was examined by defining individuals presenting with subthreshold levels of depressive symptoms, the idea being that distinguishing between depressed patients and subthreshold individuals would better represent a situation that GPs would face in primary care relative to distinguishing between depressed patients and healthy individuals. A disease prediction model of MDD was developed by comparing subthreshold individuals who did not develop MDD (reference group) against first-episode MDD patients, based on their proteomic, sociodemographic and clinical profiles. As the reference group was more similar to the patient group, this provided an additional challenge to model selection. Several methods were implemented to limit model overfitting and ensure model generalisability, and in the presence of model selection uncertainty, the multimodel approach (feature extraction and model averaging), which was developed and refined in **Chapter 3**, was applied across a set of candidate models to obtain a weighted average prediction model. To investigate the prediction of future MDD onset, this model was then

extrapolated to differentiate between subthreshold individuals who developed and did not develop MDD.

In the work reported in this chapter, blood samples were prepared by Nitin Rustogi and Rhian Lauren Preece, and run on the MS by Dr. Sureyya Ozcan. Raw MS data were processed by Dr. Sureyya Ozcan using Skyline. I was provided with the peak area values of the endogenous and SIL peptide-transitions and conducted all of the data pre-processing and analysis myself using R.

4.2. Materials and Methods

4.2.1. Clinical samples

The study presented in this chapter investigated participants from the Netherlands Study of Depression and Anxiety (NESDA), a naturalistic, longitudinal study in which 2,981 participants (aged 18-65 years) were recruited between 2004 and 2007 and followed up for up to eight years [354]. In order to reflect the various settings and developmental stages of psychopathology, participants were recruited from the general population, general practices, and mental health organisations [354]. The aims of the NESDA were: 1) to examine the long-term course and consequences of depressive and anxiety disorders, and 2) to investigate predictors of the long-term course and consequences within an epidemiological approach [354]. The protocol of the present study was approved by all relevant ethical committees (the Ethical Review Board of the VU University Medical Centre and by the local ethical review boards at the participating centres of the Leiden University Medical Centre and the Groningen University Medical Centre), written informed consent was obtained from all participants [354], and the study was conducted according to the principles of the Declaration of Helsinki [298]. Diagnoses of MDD and other psychiatric disorders were determined at the baseline and follow-up assessments using the WHO World Mental Health CIDI, version 2.1 [25]. A detailed description of the NESDA can be found in *Penninx et al. (2008)* [354].

4.2.2. Study design

For the purpose of this study, 209 participants were selected based on their disease status at the baseline and second- and fourth-year follow-up assessments, and baseline data of the 30-item Inventory of Depressive Symptomatology (IDS_{30}; self-report) [355], which measures the severity of depressive symptoms in the past seven days on a scale of zero (none) to three (severe). Using 16 items of the IDS_{30} corresponding to nine diagnostic symptoms that comprise the DSM-5 MDD criteria (and also the shortened version of the IDS_{30}, the 16-item QIDS [356]), 'subthreshold depression' at baseline was defined as presenting with two or more depressive symptoms, including at least one of sadness or anhedonia (*i.e.*, two core symptoms of the DSM), whereby a symptom was considered as present if any one of the corresponding IDS_{30} items was above zero (**Table 4.1**).

To identify early biomarkers or indicators of MDD, it would be ideal to test for differences between subthreshold individuals who later developed and did not develop MDD. However, as the number of subthreshold individuals who developed MDD was limited, the model was first trained to differentiate between 86 subthreshold individuals who had no current or lifetime diagnosis of MDD at the baseline assessment and did not develop MDD by the fourth-year follow-up assessment (reference group) and 86 recent-onset MDD patients who experienced their first and only MDE within a month before the baseline assessment (training set patient group). To provide a fair comparison, recent-onset MDD patients also had to fulfil the criteria for baseline subthreshold depression. The model was subsequently extrapolated to predict the probability of developing MDD in the shared reference group and 37 subthreshold individuals who had no current or lifetime diagnosis of MDD at the baseline assessment and developed MDD by the second-year (n = 21) or fourth-year (n = 16) follow-up assessment (extrapolation test set patient group).

None of the selected participants were diagnosed with BD, obsessive compulsive disorder, severe substance use disorder or psychotic disorder at the baseline assessment, and/or with BD at the follow-up assessments. Comorbid anxiety disorder was not used as an exclusion criterion.

Table 4.1. Nine DSM-5 depressive symptoms and corresponding IDS_{30} items.

For a diagnosis of MDD under the DSM-5, five or more of the nine symptoms need to be present, including at least one core symptom (*) of depressed mood or anhedonia, during the same two-week period and represent a change from previous functioning [1]. The present study defined subthreshold symptomatic individuals at baseline using 16 items of the IDS_{30} [355] that correspond to the nine DSM symptom domains (and comprise the shortened version of the IDS_{30}, the 16-item QIDS [356]): individuals had to present two or more depressive symptoms including one core symptom, whereby any one of the corresponding items had to be above zero for a symptom to be considered as present. Abbreviations: DSM (Diagnostic and Statistical Manual of Mental Disorders); IDS (Inventory of Depressive Symptomatology); MDD (major depressive disorder); QIDS (Quick Inventory of Depressive Symptomatology).

DSM-5 symptom	IDS item	IDS description
*Depressed mood	IDS 5	Sadness
*Anhedonia	IDS 19	General interest
Decrease or increase in weight or appetite	IDS 11	Decreased appetite
	IDS 12	Increased appetite
	IDS 13	Decreased weight
	IDS 14	Increased weight
Insomnia or hypersomnia	IDS 1	Early insomnia
	IDS 2	Middle insomnia
	IDS 3	Late insomnia
	IDS 4	Hypersomnia
Psychomotor agitation or retardation	IDS 23	Psychomotor retardation
	IDS 24	Psychomotor agitation
Fatigue or loss of energy	IDS 20	Energy level
Worthlessness or inappropriate guilt	IDS 16	Self-criticism
Diminished ability to concentrate or indecisiveness	IDS 15	Concentration/decision-making
Recurrent thoughts of death or suicide	IDS 18	Thoughts of death or suicide

4.2.3. Sociodemographic and clinical characteristics

Sociodemographic and lifestyle information of the participants was collected at the NESDA baseline assessment [354]. This included sex, age, BMI, education level, physical activity, smoking, alcohol abuse, recreational drug use, employment status, family history, childhood trauma, chronic diseases, and medication use (**Table 4.2**). Clinical features were derived from the baseline IDS_{30} data: 28 depressive symptoms were derived from 30 IDS_{30} items (after items on increase or decrease in weight and appetite were aggregated into single domains of weight/appetite increase or decrease), and the IDS_{30} total score and severity classification were determined.

The sociodemographic and lifestyle characteristics of participants in different clinical groups were compared using Kruskal-Wallis *H* tests for numerical variables and χ^2 tests for categorical variables (see **Section 2.5.6**). *P*-values < 0.05 were considered to be statistically significant (**Table 4.2**).

4.2.4. Targeted protein quantification

4.2.4.1. Serum sample preparation

Blood serum samples of the participants were collected at the NESDA baseline assessment [354]. Two hundred and nine serum samples of the selected participants were randomised to allocate equal numbers of reference and patient groups, training and test sets, and males and females across four experimental plates for LC-MS/MS analysis. Samples were and prepared in a 96-well plate format as described in **Section 2.3.1.1.**

4.2.4.2. LC-MS/MS analysis

One hundred and forty-six peptides representing 77 proteins, the majority previously associated with psychiatric disorders, were measured using targeted MRM-MS analysis. Peptides were quantified at the transition level. More details on LC-MS/MS analysis can be found in **Section 2.3.2.** The researchers conducting the sample preparation and the MS analysis were blinded to the clinical status of the participants.

4.2.4.3. QC samples

Pooled and Sigma serum QC samples were used to assess the technical variation associated with instrument performance and sample preparation as described in **Section 2.3.3**. A pooled QC sample, which was prepared by pooling together the digested clinical serum samples, was injected once a day along with the clinical samples for the duration of the entire study (28 injections in total). Twenty-four Sigma serum QC samples were prepared following the same protocol as the clinical samples and distributed across the experimental plates. More details on targeted protein quantification can be found in **Section 2.3**.

4.2.5. MS data pre-processing

Pre-processing of the MS data was carried out as described in **Chapter 2** (see **Section 2.4**). Quantifier transitions were selected for each peptide based on the abundance values of the endogenous and the SIL peptide-transitions. A quantifier transition was selected as the transition that was consistently most abundant in at least 80% of the sample runs in both the endogenous and SIL peptides. Peptides for which quantifier transitions could not be selected based on these criteria, which can occur in the presence of competing transitions, the raw peak data were examined for manual selection. Subsequent analyses were conducted using data of the quantifier transitions only. Peptide quantification was based on the relative abundance of the endogenous and SIL peptide quantifier transitions, and the abundance ratio was $\log_2$-transformed for statistical analysis.

4.2.6. Statistical quality control

PCA was conducted to identify any outliers based on the $\log_2$-transformed abundance ratios of 146 peptides (**Figure 4.1**). No outlier samples were identified.

In addition, the geometric CV was used to estimate the magnitude of variation in the proteomic data (based on the abundance ratios of 147 peptides), as described in **Chapter 2** (see **Section 2.4.6**). The technical variation was assessed by calculating the geometric CVs of peptide abundance ratios of the pooled (variation in MS instrument performance) and the Sigma serum (variation in sample preparation) QC samples across MS runs. The median CV values for the pooled and Sigma serum QC runs were 6.7% and 22.9%, respectively. The biological variation was assessed by calculating the geometric CVs of peptide abundance ratios of clinical samples within each clinical sample group. The median CV values were 59.2%, 51.7% and 49.7% for

the training set patient group, extrapolation test set patient group and shared reference group, respectively.

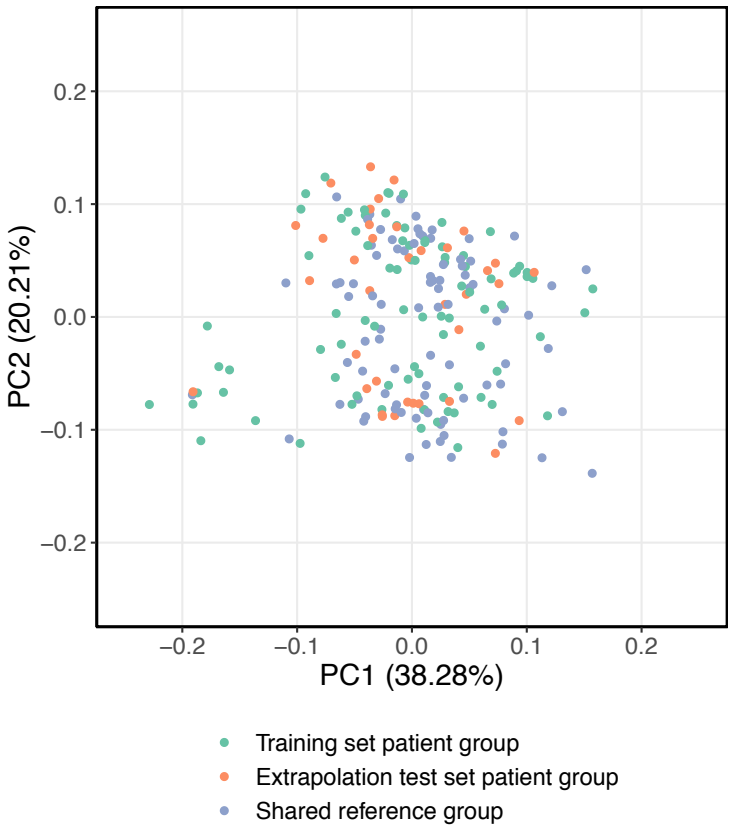


Figure 4.1. PCA plot of protein peptide abundance ratios of clinical samples.
The scores for the first two PCs are plotted with the percentage of variation accounted for by each PC shown in the axis labels. Data points are coloured according to the allocation of clinical samples into the training set patient group (first-episode MDD patients), the extrapolation test set patient group (subthreshold symptomatic individuals who developed MDD within two or four years) and the shared reference group (subthreshold symptomatic individuals who did not develop MDD within four years). Abbreviations: MDD (major depressive disorder); PC (principal component).

4.2.7. Missing data

There were no peptides with missing values, and no sociodemographic or clinical variables with more than 5% missing values. Missing values in sociodemographic and clinical variables were replaced using multiple imputation. Methods used for imputation were: predictive mean matching for numeric data, logistic regression imputation for binary data (factor with two

levels), and multinomial logistic regression imputation for categorical data (factor with more than two levels). The final complete dataset was produced by finding the central estimate of the missing values across five imputed datasets, using the median and the mode for numeric and categorical variables, respectively. Multiple imputation was conducted using the R package mice [357].

4.2.8. Data analysis

4.2.8.1. Model selection

A total of 198 features (146 proteomic, 22 sociodemographic and 30 clinical) was analysed for model selection to predict the probability of MDD outcome in the training set. The full list of the analysed proteins and peptides can be found in the **Appendix** (**Table A. 1**). Categorical variables were represented as sets of dummy variables. Group LASSO regression was used to allow sets of dummy variables derived from categorical variables to be selected together. Ten-fold cross-validation was used to further reduce overfitting by selecting the value of the shrinkage parameter λ that resulted in the most regularised model. More details on group LASSO regression and *K*-fold cross-validation can be found in **Chapter 2** (see **Sections 2.5.3 and 2.5.4)**.

One hundred models were generated by repeatedly applying group LASSO regression with ten-fold cross-validation on the training set. As shown in **Chapter 3**, this approach allowed for model selection uncertainty to be investigated by evaluating the sensitivity of model selection to small changes in the data that resulted from the random partitioning in ten-fold cross-validation. Relative model weights and relative importance of the features were measured as described in **Chapter 3** (see **Section 3.2.5**). Briefly, the AICc was used to calculate the Akaike weight of each model, which was a value between 0 and 1 and interpreted as the probability that the model was the best approximating model for the data [284]. The sum of weights of all models was equal to 1. The relative importance of each feature was assessed by measuring the selection fraction (*i.e.*, the proportion of models out of 100 in which it was selected), which was a value between 0 and 1. Moreover, models which comprised of the same combinations of features selected were grouped together, and the relative importance of subsets of features occurring together was quantified by measuring the frequency of occurrence of each model

group, as well as by summing the Akaike weights of all corresponding models to estimate the probability that the selected combination of features comprised the best approximating model.

4.2.8.2. Feature extraction and model averaging

As discussed in **Chapter 3**, when there was one strongly supported model group (*e.g.*, $w > 0.9$), inference and prediction could be based on that model alone [284]. In this instance, the coefficients of the features in the superior model group were estimated by averaging over the corresponding set of models.

However, in the absence of a superior model group, that is, when there was uncertainty in model selection, the multimodel approach that was developed in **Chapter 3** was applied to obtain more reproducible predictions of the probability of MDD outcome. Based on the findings in **Chapter 3**, feature extraction was applied using a selection fraction threshold of 0.9, such that only features with selection fractions ≥ 0.9 were included in the prediction model to limit overfitting. This was followed by model averaging across all 100 models to obtain better estimates of feature coefficients [284], [289], [290]. The weighted average coefficient of a given feature was estimated across the 100 models as described in **Chapter 3** (see **Section 3.2.5.7**).

4.2.8.3. Predictive performance

Predictive performance of the models when applied to the training and test sets was evaluated by plotting ROC curves and measuring the AUC. More details on model predictive performance can be found in **Chapter 2** (see **Section 2.5.5)**.

4.3. Results

One hundred and forty-six proteomic, 22 sociodemographic and 30 clinical features (198 total) were measured in the training set patient group of 86 first-episode MDD patients, the extrapolation test set patient group of 37 subthreshold symptomatic individuals who developed MDD within two or four years, and the shared reference group of 86 subthreshold symptomatic individuals who did not develop MDD within four years (**Table 4.2**).

Table 4.2. Sociodemographic and lifestyle characteristics of individuals in the training set patient group (first-episode MDD patients), the extrapolation test set patient group (subthreshold symptomatic individuals who developed MDD within two or four years) and the shared reference group (subthreshold symptomatic individuals who did not develop MDD within four years).
Numerical variables are shown as the mean (standard deviation), and binary and categorical variables are shown as the percentage of participants in each category. Abbreviations: BMI (body mass index); IDS (Inventory of Depressive Symptomatology); MDD (major depressive disorder).

	Shared reference group	**Training set patient group**	**Extrapolation test set patient group**	***p*-value**
n	86	86	37	
Sex % (male/female)	35/65	48/52	32/68	0.1399
Age (years)	37.8 (14.1)	41.8 (12.2)	38.5 (14)	0.1421
Body mass index (kg/m^2)	23.8 (4.4)	26.7 (6)	25.6 (5)	0.0018
Education % (basic/intermediate/high)	8/42/50	6/67/27	3/59/38	0.0137
Physical activity % (low/moderate/high)	23/48/29	30/44/26	27/41/32	0.8157
Smoking % (yes/no)	31/69	38/62	32/68	0.6041
Alcohol abuse % (yes/no)	21/79	40/60	19/81	0.0097
Weekly alcohol consumption (number of drinks per week)	8 (11)	6.8 (11)	6.3 (7.4)	0.1615
Recreational drug use (past month) % (yes/no)	7/93	8/92	8/92	0.9535
Partner % (yes/no)	69/31	57/43	70/30	0.1952
Children % (yes/no)	44/56	48/52	51/49	0.7522

Employment % (employed/unemployed/retired/ occupationally disabled)	77/17/2/3	60/17/2/20	65/30/0/5	0.0097
Absent from work due to health problems (past 6 months) % (yes/no/not applicable)	43/35/22	41/21/38	35/32/32	0.1209
Childhood life event index score	0.3 (0.5)	0.2 (0.5)	0.2 (0.5)	0.8888
Childhood trauma index score	0.5 (0.9)	1 (1.3)	0.8 (1.1)	0.0086
Number of negative life events (past year)	1 (1)	1 (1.1)	0.9 (1)	0.9182
Family history % (yes/no)	73/27	87/13	84/16	0.0591
Heart disease % (yes/no)	1/99	5/95	5/95	0.3323
Diabetes % (yes/no)	3/97	6/94	11/89	0.2774
Other chronic disease % (yes/no)	26/74	40/60	22/78	0.0592
Anti-inflammatory drug % (yes/no)	2/98	8/92	8/92	0.2069
Heart medication % (yes/no)	10/90	23/77	8/92	0.027
IDS_{30} total score	14.9 (7.4)	37.4 (11.5)	20.2 (8.8)	< 0.0001

4.3.1. Analysis 1: model selection including IDS_{30} total score

When all 198 features were used, there was minimal uncertainty in model selection. The number of features selected in a model ranged from one to six with an average of one (**Figure 4.2a**). IDS_{30} total score was selected 100 times, one peptide was selected four times, and four peptides were selected once; the remaining features were never selected. Three model groups were identified based on the combination of features selected (**Table 4.3**). The most frequently occurring model group consisting of the IDS_{30} total score alone occurred 96 times and had a model probability of 0.98. Given the strong support for this model group (Model 1), there was no need for feature extraction, and the average feature coefficient was estimated using the corresponding 96 models (**Table 4.4**).

The resulting single-feature model of IDS_{30} total score showed an excellent predictive performance when applied to the training set (AUC = 0.95; **Figure 4.3a**), and a poor performance when extrapolated to the test set (AUC = 0.68; **Figure 4.3b**). This suggests that while first-episode MDD patients could be accurately distinguished from subthreshold

individuals who did not develop MDD based on the IDS_{30} total score alone, the differentiation was much more difficult between subthreshold individuals who developed and did not develop MDD as both groups had minimal symptoms resulting in more similar scores (**Table 4.2**).

4.3.2. Analysis 2: model selection excluding IDS_{30} total score

To improve the model predictive performance on the extrapolation test set, the analysis was repeated after excluding IDS_{30} total score to allow for other features to be selected (197 features). In this case, the number of features selected in a model ranged from ten to 27, with a median of 14, demonstrating a degree of uncertainty in model selection (**Figure 4.2b**). Twelve features were selected at least 90 times out of 100, among which nine features were selected 100 times. Twenty-eight features were selected at least once, and 169 features were never selected. Seventeen model groups were identified based on the combination of features selected together (**Table 4.3**). Two competing models consisting of 13 and 14 features occurred most frequently, 30 and 29 times, and had model probabilities of 0.22 and 0.50, respectively; the 13 features were a subset of the 14 features. This demonstrates that the frequency of occurrence did not necessarily correspond to the probability of being the best approximating model for the given data. The remaining model groups each occurred eight times or fewer. As there was considerable variability in feature selection and no strongly supported model group, feature extraction and model averaging were implemented across all 100 models.

The resulting average model (Model 2) was comprised of 12 features that had selection fractions greater than 0.9 (**Table 4.4**). Six peptides representing six proteins (α-1-antichymotrypsin (AACT), apolipoprotein E (APOE), apolipoprotein H (APOH), fetuin-A (FETUA), haemoglobin subunit α (HBA) and glycoprotein phospholipase D (PHLD)) were included, as well as three sociodemographic factors (BMI, childhood trauma and education level), and three depressive symptoms (sadness, fatigue and leaden paralysis). The 12-feature average prediction model showed an excellent predictive performance when applied to the training set (AUC = 0.94; **Figure 4.3a**), and a fair predictive performance when extrapolated to the test set (AUC = 0.75; **Figure 4.3b**). Here, the reduced performance on the latter can be explained by subthreshold individuals who developed MDD generally displaying weaker indications of disease (*i.e.*, more similar to the reference group) compared to first-episode MDD patients (**Figure 4.4**), as expected.

In order to assess the relative contributions of the selected proteomic biomarkers and non-biological features in predicting MDD, a prediction model consisting only of the six protein peptides and another model consisting only of the three sociodemographic factors and the three depressive symptoms were applied to the data sets. A biomarker-only model achieved a good predictive performance on the training set (AUC = 0.76), and a fair predictive performance on the extrapolation test set (AUC = 0.70). On the other hand, a sociodemographic-clinical-only model achieved an excellent predictive performance on the training set (AUC = 0.91), and a poor predictive performance on the extrapolation test set (AUC = 0.67). This demonstrates that combining the biomarkers and the non-biological features resulted in improved predictive performance on all data sets, and importantly, in predicting future onset of MDD among subthreshold individuals on the extrapolation test set.

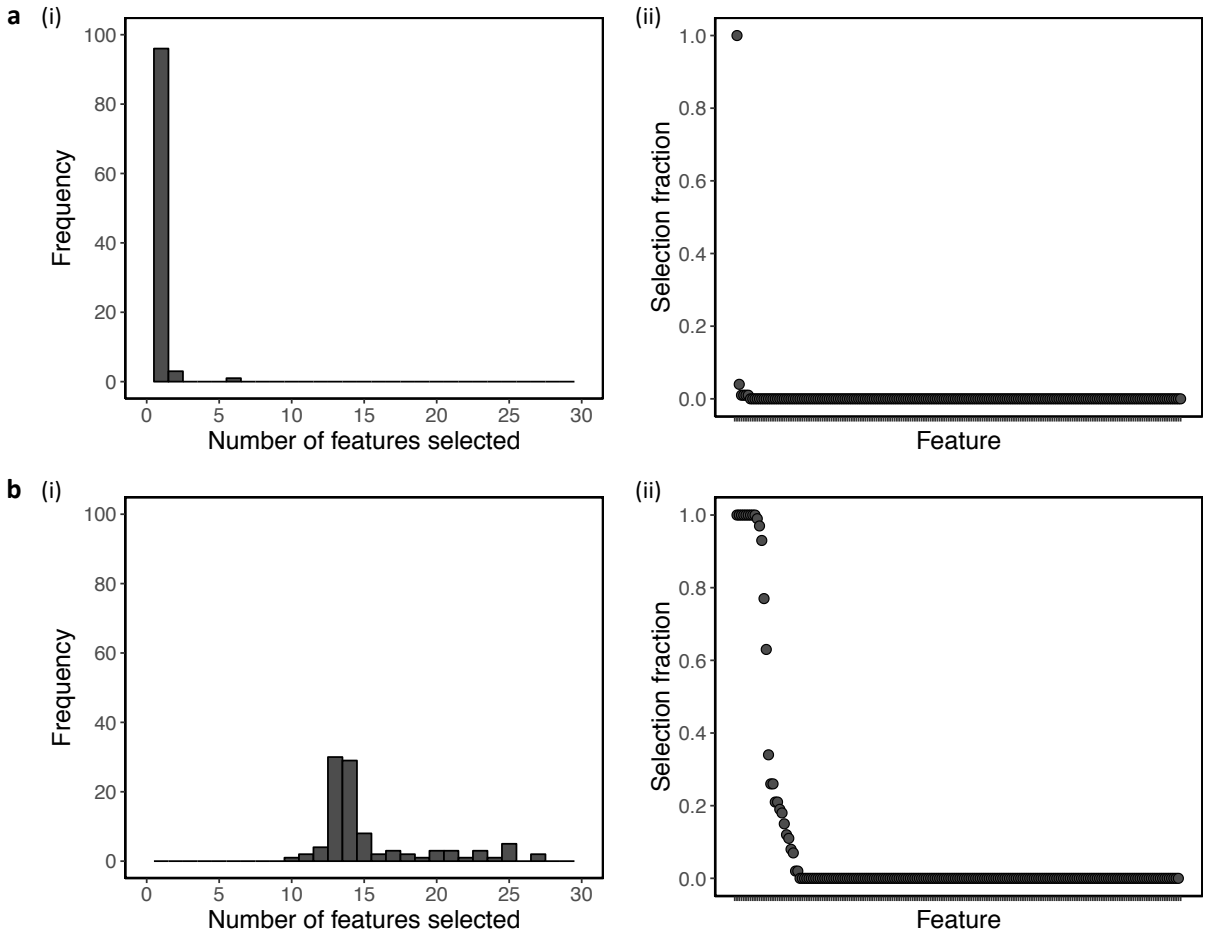


Figure 4.2. Feature selection across 100 models obtained from repeated application of group LASSO regression with ten-fold cross-validation on the training set.

(a) Analysis 1: model selection including IDS_{30} total score (198 features). (b) Analysis 2: model selection excluding IDS_{30} total score (197 features). (i) Number of features selected in each model. (ii) Selection fractions of each feature. Abbreviations: IDS (Inventory of Depressive Symptomatology; LASSO (least absolute shrinkage and selection operator).

Table 4.3. Model groups based on the combinations of features selected.

Three model groups were identified in Analysis 1 (model selection including IDS_{30} total score), and 17 model groups were identified in Analysis 2 (model selection excluding IDS_{30} total score). For each model group, the frequency of occurrence and the model probability (summed Akaike weight) are reported. Model groups are shown in descending order of frequency of occurrence. Proteomic features are represented in a protein_peptide format, and their abbreviations can be found in the **Appendix** (**Table A. 1**).

Model group	Number of features	Frequency	Model probability
Analysis 1			
IDS_{30} total score	1	96	0.9825
PHLD_NQVVIAAGR + IDS_{30} total score	2	3	0.0171
FETUA_HTLNQIDEVK + IGHG2_GLPAPIEK + PHLD_NQVVIAAGR + RET4_QEELCLAR + SHBG_LPLVPALDGCLR + IDS_{30} total score	6	1	0.0004
Analysis 2			
AACT_ADLSGITGAR + AACT_EIGELYLPK + APOE_ALMDETMK + APOH_EHSSLAFWK + FETUA_HTLNQIDEVK + HBA_MFLSFPTTK + PHLD_NQVVIAAGR + Education + BMI + Childhood trauma + Sadness + Fatigue + Leaden paralysis	13	30	0.2169
A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + AACT_EIGELYLPK + APOE_ALMDETMK + APOH_EHSSLAFWK + FETUA_HTLNQIDEVK + HBA_MFLSFPTTK + PHLD_NQVVIAAGR + Education + BMI + Childhood trauma + Sadness + Fatigue + Leaden paralysis	14	29	0.5024
A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + APOE_ALMDETMK + APOH_EHSSLAFWK + APOL1_VNEPSILEMSR + CO8A_MESLGITSR + FETUA_HTLNQIDEVK + HBA_FLASVSTVLTSK + HBA_MFLSFPTTK +	23	3	< 0.0001

HRG_ADLFYDVEALDLESPK + IGHG2_GLPAPIEK + ITIH4_GPDVLTATVSGK + PHLD_NQVVIAAGR + Education + BMI + Childhood life event + Childhood trauma + Sadness + Mood reactivity + Self-criticism + Fatigue + Pleasure/enjoyment + Leaden paralysis			
A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + APOE_ALMDETMK + APOH_EHSSLAFWK + APOL1_VNEPSILEMSR + CO6_SEYGAALAWEK + CO8A_MESLGITSR + FETUA_HTLNQIDEVK + HBA_FLASVSTVLTSK + HBA_MFLSFPTTK + HRG_ADLFYDVEALDLESPK + IGHG2_GLPAPIEK + ITIH4_GPDVLTATVSGK + PHLD_NQVVIAAGR + Education + BMI + Childhood life event + Childhood trauma + Sadness + Mood reactivity + Mood quality + Self-criticism + Fatigue + Pleasure/enjoyment + Leaden paralysis	25	5	< 0.0001
A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + AACT_EIGELYLPK + APOE_ALMDETMK + APOH_EHSSLAFWK + FETUA_HTLNQIDEVK + HBA_FLASVSTVLTSK + HBA_MFLSFPTTK + PHLD_NQVVIAAGR + Education + BMI + Childhood trauma + Sadness + Fatigue + Leaden paralysis	15	8	0.1530
AACT_ADLSGITGAR + AACT_EIGELYLPK + APOE_ALMDETMK + APOH_EHSSLAFWK + HBA_MFLSFPTTK + PHLD_NQVVIAAGR + BMI + Childhood trauma + Sadness + Fatigue + Leaden paralysis	11	2	0.0158
AACT_ADLSGITGAR + AACT_EIGELYLPK + APOE_ALMDETMK + APOH_EHSSLAFWK + FETUA_HTLNQIDEVK + HBA_MFLSFPTTK + PHLD_NQVVIAAGR + BMI + Childhood trauma + Sadness + Fatigue + Leaden paralysis	12	4	0.0963

A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + AACT_EIGELYLPK + APOE_ALMDETMK + APOH_EHSSLAFWK + FETUA_HTLNQIDEVK + HBA_FLASVSTVLTSK + HBA_MFLSFPTTK + IGHG2_GLPAPIEK + PHLD_NQVVIAAGR + Education + BMI + Childhood trauma + Sadness + Self-criticism + Fatigue + Leaden paralysis	17	3	0.0002
A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + APOE_ALMDETMK + APOH_EHSSLAFWK + CO8A_MESLGITSR + FETUA_HTLNQIDEVK + HBA_FLASVSTVLTSK + HBA_MFLSFPTTK + HRG_ADLFYDVEALDLESPK + IGHG2_GLPAPIEK + ITIH4_GPDVLTATVSGK + PHLD_NQVVIAAGR + Education + BMI + Childhood trauma + Sadness + Mood reactivity + Self-criticism + Fatigue + Pleasure/enjoyment + Leaden paralysis	21	3	< 0.0001
A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + APOE_ALMDETMK + APOH_EHSSLAFWK + CO8A_MESLGITSR + FETUA_HTLNQIDEVK + HBA_FLASVSTVLTSK + HBA_MFLSFPTTK + HRG_ADLFYDVEALDLESPK + IGHG2_GLPAPIEK + PHLD_NQVVIAAGR + Education + BMI + Childhood trauma + Sadness + Self-criticism + Fatigue + Leaden paralysis	18	2	< 0.0001
A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + APOE_ALMDETMK + APOH_EHSSLAFWK + APOL1_VNEPSILEMSR + CO6_SEYGAALAWEK + CO8A_MESLGITSR + FCN3_YGIDWASGR + FETUA_HTLNQIDEVK + HBA_FLASVSTVLTSK + HBA_MFLSFPTTK + HRG_ADLFYDVEALDLESPK + IGHG2_GLPAPIEK + ITIH4_GPDVLTATVSGK +	27	2	< 0.0001

PHLD_NQVVIAAGR + Education + BMI + Childhood life event + Childhood trauma + Sadness + Mood reactivity + Mood quality + Self-criticism + Suicidal thoughts + Fatigue + Pleasure/enjoyment + Leaden paralysis			
A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + APOE_ALMDETMK + APOH_EHSSLAFWK + APOL1_VNEPSILEMSR + CO6_SEYGAALAWEK + CO8A_MESLGITSR + FETUA_HTLNQIDEVK + HBA_FLASVSTVLTSK + HBA_MFLSFPTTK + HRG_ADLFYDVEALDLESPK + IGHG2_GLPAPIEK + ITIH4_GPDVLTATVSGK + PHLD_NQVVIAAGR + Education + BMI + Childhood life event + Childhood trauma + Sadness + Mood reactivity + Self-criticism + Fatigue + Pleasure/enjoyment + Leaden paralysis	24	1	< 0.0001
A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + APOE_ALMDETMK + APOH_EHSSLAFWK + FETUA_HTLNQIDEVK + HBA_FLASVSTVLTSK + HBA_MFLSFPTTK + IGHG2_GLPAPIEK + PHLD_NQVVIAAGR + Education + BMI + Childhood trauma + Sadness + Self-criticism + Fatigue + Leaden paralysis	16	2	0.0006
AACT_ADLSGITGAR + AACT_EIGELYLPK + APOH_EHSSLAFWK + HBA_MFLSFPTTK + PHLD_NQVVIAAGR + BMI + Childhood trauma + Sadness + Fatigue + Leaden paralysis	10	1	0.0148
A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + APOE_ALMDETMK + APOH_EHSSLAFWK + CO8A_MESLGITSR + FETUA_HTLNQIDEVK + HBA_FLASVSTVLTSK + HBA_MFLSFPTTK + HRG_ADLFYDVEALDLESPK + IGHG2_GLPAPIEK + ITIH4_GPDVLTATVSGK + PHLD_NQVVIAAGR + Education + BMI +	20	3	< 0.0001

Childhood trauma + Sadness + Mood reactivity + Self-criticism + Fatigue + Leaden paralysis			
A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + APOE_ALMDETMK + APOH_EHSSLAFWK + APOL1_VNEPSILEMSR + CO8A_MESLGITSR + FETUA_HTLNQIDEVK + HBA_FLASVSTVLTSK + HBA_MFLSFPTTK + HRG_ADLFYDVEALDLESPK + IGHG2_GLPAPIEK + ITIH4_GPDVLTATVSGK + PHLD_NQVVIAAGR + Education + BMI + Childhood trauma + Sadness + Mood reactivity + Self-criticism + Fatigue + Pleasure/enjoyment + Leaden paralysis	22	1	< 0.0001
A2MG_NEDSLVFVQTDK + AACT_ADLSGITGAR + APOE_ALMDETMK + APOH_EHSSLAFWK + CO8A_MESLGITSR + FETUA_HTLNQIDEVK + HBA_FLASVSTVLTSK + HBA_MFLSFPTTK + HRG_ADLFYDVEALDLESPK + IGHG2_GLPAPIEK + PHLD_NQVVIAAGR + Education + BMI + Childhood trauma + Sadness + Mood reactivity + Self-criticism + Fatigue + Leaden paralysis	19	1	< 0.0001

Table 4.4. Features included in the two prediction models from Analysis 1 and Analysis 2.
Model 1 (one feature) was based on the superior model group in Analysis 1 (model selection including IDS_{30} total score), and Model 2 (12 features) was developed by implementing feature extraction and model averaging in Analysis 2 (model selection excluding IDS_{30} total score) in the absence of a superior model group. The selection fraction and the weighted mean coefficient of the features are shown. Proteomic features are represented in a protein_peptide format. Categorical features (education, sadness, fatigue and leaden paralysis) are represented as sets of dummy variables. Abbreviations: AACT (α-1-antichymotrypsin); APOE (apolipoprotein E); APOH (apolipoprotein H); BMI (body mass index); FETUA (fetuin-A); HBA (haemoglobin subunit α); IDS (Inventory of Depressive Symptomatology); PHLD (glycoprotein phospholipase D).

Feature	Model	Selection fraction	Weighted mean coefficient
Proteomic			
AACT_ADLSGITGAR	2	1.00	0.122
APOE_ALMDETMK	2	0.99	-0.195
APOH_EHSSLAFWK	2	1.00	0.080
FETUA_HTLNQIDEVK	2	0.97	0.082
HBA_MFLSFPTTK	2	1.00	0.231
PHLD_NQVVIAAGR	2	1.00	0.286
Sociodemographic			
BMI	2	1.00	0.291
Childhood trauma	2	1.00	0.115
Education; Intermediate	2	0.93	0.065
Education; High	2	0.93	-0.055
Clinical			
Sadness; Mild	2	1.00	-0.681
Sadness; Moderate	2	1.00	0.819
Sadness; Severe	2	1.00	0.369
Fatigue; Mild	2	1.00	-0.124
Fatigue; Moderate	2	1.00	0.339
Fatigue; Severe	2	1.00	0.085
Leaden paralysis; Mild	2	1.00	-0.145
Leaden paralysis; Moderate	2	1.00	0.219
Leaden paralysis; Severe	2	1.00	0.272
IDS_{30} total score	1	1.00	0.346

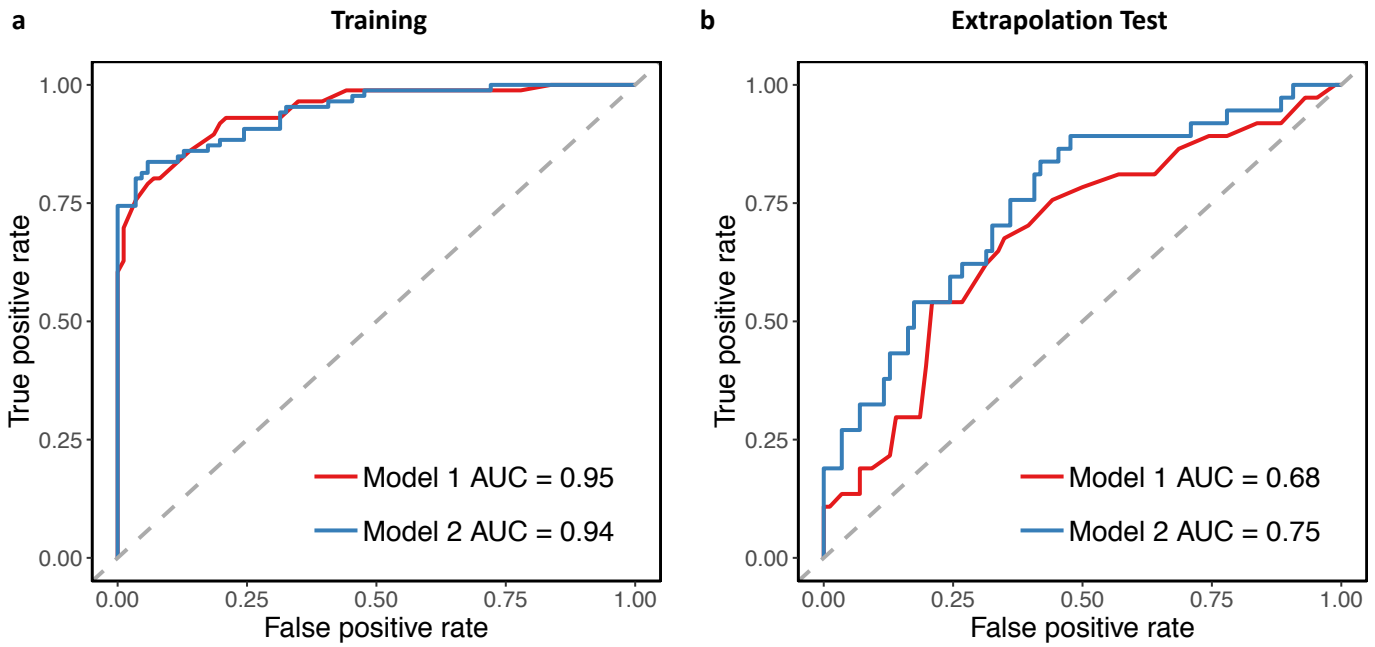


Figure 4.3. ROC curves showing model performance in predicting the probability of MDD outcome.

Model 1 consisted of IDS_{30} total score and Model 2 consisted of six proteins (α-1-antichymotrypsin, apolipoprotein E, apolipoprotein H, fetuin-A, haemoglobin subunit α and glycoprotein phospholipase D), three sociodemographic factors (BMI, childhood trauma and education level), and three depressive symptoms (sadness, fatigue and leaden paralysis). The prediction models were applied to predict the probability of MDD outcome in: (a) the training set (86 first-episode MDD patients *vs* 86 subthreshold individuals who did not develop MDD within four years), and (b) the extrapolation test set (37 subthreshold individuals who developed MDD within two or four years *vs* 86 subthreshold individuals who did not develop MDD within four years). Abbreviations: AUC (area under the curve); IDS (Inventory of Depressive Symptomatology); MDD (major depressive disorder); ROC (receiver operator characteristic).

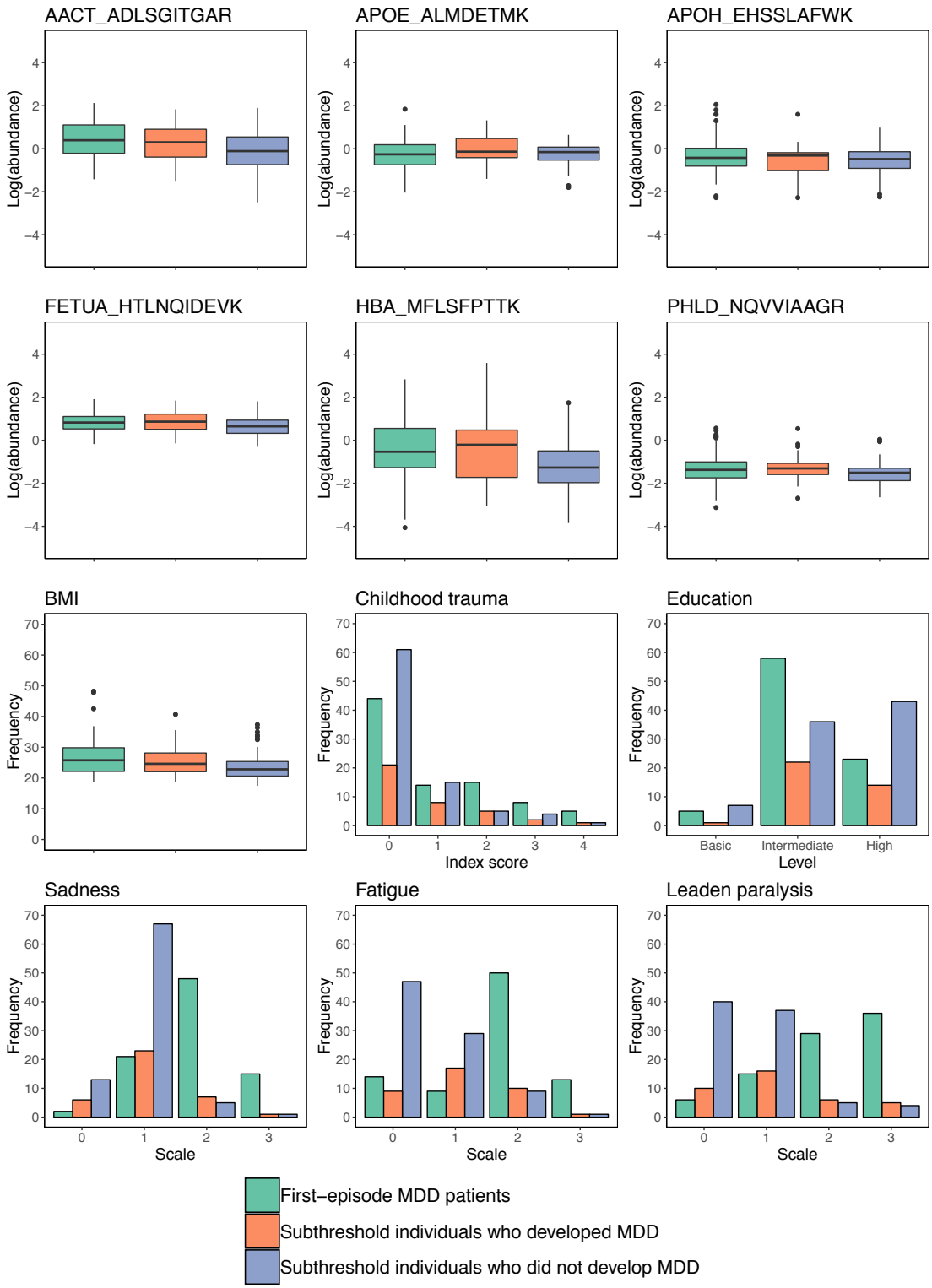


Figure 4.4. Disease indications of MDD represented by 12 features comprising Model 2.

The distribution of data for individuals in the training set patient group (first-episode MDD patients), the extrapolation test set patient group (subthreshold symptomatic individuals who developed MDD within two or four years), and the shared reference group (subthreshold symptomatic individuals who did not develop MDD within four years) is shown. Protein abundances are represented by the $\log_2$-

transformed peptide abundance ratios. The severity of depressive symptoms is represented on scale of zero (none) to three (severe). Numeric features are illustrated using boxplots, and categorical features are illustrated using bar charts.

4.4. Discussion

This chapter evaluated the accuracy with which future onset of MDD could be predicted in subthreshold symptomatic individuals by extrapolating a disease prediction model of MDD that was trained to differentiate between first-episode MDD patients and subthreshold symptomatic individuals who did not develop MDD. A fair predictive performance (AUC = 0.75) was obtained, which was promising given that it involved going beyond the scope of the model, although there is potential for improvement. Consequently, this study demonstrated that some indicators of future MDD onset could be detected in subthreshold individuals using samples and data collected up to four years prior to diagnosis. This has important clinical implications with regards to enabling healthcare professionals to identify individuals with high probabilities of incident MDD and subsequently provide appropriate early intervention strategies, in light of accumulating evidence that subthreshold individuals have an increased risk of developing MDD [67], [72], [76] and that subthreshold depression may represent a prodromal stage of MDD [55], [358].

The disease prediction model was comprised of a combination of 12 proteomic, sociodemographic and clinical features. Six proteins (represented by six peptides) were identified as biomarkers of MDD: AACT, APOE, APOH, FETUA, HBA and PHLD, which have functional roles in acute-phase response, lipid transport and metabolism, blood coagulation and oxygen transport [359]. MDD was associated with increased serum levels of AACT, APOH, FETUA, HBA and PHLD, and decreased serum levels of APOE. Alterations in peripheral proteins involved in inflammatory response, the HPA axis, and carbohydrate and lipid metabolism have been reported in previous biomarker studies on depression [14], [222]–[227] (see **Chapter 1**). In particular, the finding of increased levels of acute-phase proteins, AACT and FETUA, is consistent with previous reports of increased blood levels of acute-phase proteins in MDD [141], [150], [151], although the protein AACT itself has not previosuly been found to be associated with the disorder. Nevertheless, it has been reported that cortisol resulted

in an increase in the mRNA levels of SERPINA3, the gene encoding the AACT protein, demonstrating the pro-inflammatory properties of glucocorticoids and thereby providing a potential mechanism by which stress can lead to increased inflammation [360], [361]. Plasma FETUA has been identified as a biomarker of depression in another recent study; however, the reported direction of change was inconsistent with that reported in the present study (*i.e.*, depression was found to be associated with reduced plasma levels of FETUA) [362]. Moreover, evidence from genetic studies suggests that the presence of the APOE ε4 allele may enhance the risk of developing MDD by increasing susceptibility to the adverse effects of stress [363], [364]. While the association of APOH with MDD has not previously been reported, other apolipoproteins (apolipoprotein A, apolipoprotein B, apolipoprotein C-III and apolipoprotein D) have been linked to the disorder, with altered serum levels found in MDD patients compared to healthy controls [223], [239], [241], [243], [365].

Three sociodemographic factors (BMI, childhood trauma and education level) and three depressive symptoms (sadness, fatigue and leaden paralysis) were identified as important predictors of MDD outcome. The relationship between BMI and depression is well known [366], [367], and some studies have reported a shared pathophysiology between obesity and depression, including dysregulation of the HPA axis and inflammatory response [130]–[132]. MDD was found to be associated with a higher childhood trauma index score, which measured experiences of emotional neglect, psychological abuse, physical abuse and sexual abuse in early life. Consistent with this, adverse or traumatic experience in childhood has been found as a strong risk factor for developing depression in adulthood [115]–[117], and this relationship has been suggested to be reflected in disturbances in the neuroendocrine and autoimmune stress response system [146], [147]. The link between education level and depression is less well established, with some studies reporting a decreased risk and others reporting an increased risk of depression with a higher education level [368]–[370]. Moreover, the identification of depressive symptoms as key predictors of MDD supports the idea that individual symptoms are associated with different risk factors, and that they are not interchangeable as assumed by the current diagnostic approach in which symptoms are added together [371], [372]. Note that sadness is one of the two core symptoms of depression according to both the DSM-5 and ICD-11 and was also required as a core symptom for the definition of subthreshold depression in this study. Fatigue (reduced energy level) is specified as a core symptom in the ICD-11 but not in the DSM-5. The identification of leaden paralysis (heaviness in arms and legs) as a key feature was interesting, as it is a symptom of atypical depression, a subtype of depression, and

not included in the DSM-5 or ICD-11 criteria for general MDD. Overall, this study demonstrated the advantage of integrating different aspects of patient data (*i.e.*, proteomic, sociodemographic and clinical) for developing a clinically useful disease prediction model. The combined biomarker-sociodemographic-clinical prediction model resulted in an improved predictive performance for predicting future MDD onset in subthreshold symptomatic individuals in relation to both the biomarker-only and sociodemographic-clinical-only models.

Furthermore, this study showed that the combined use of feature extraction and model averaging could effectively address model selection uncertainty and result in a parsimonious prediction model of MDD. This builds upon the work in **Chapter 3** and further demonstrates the utility of the multimodel approach in producing reproducible predictions of a complex psychiatric disorder, not only when applied to investigate individuals with distinct clinical conditions (*e.g.*, first-onset drug-naïve schizophrenia patients *vs* healthy controls, as in **Chapter 3**), but also when applied to investigate more clinically relevant individuals with less distinct conditions (*e.g.*, first-episode MDD patients *vs* subthreshold individuals, as in this chapter). In comparison to the one-feature model of IDS_{30} total score (Model 1) that was based on the superior model group in Analysis 1, the 12-feature model (Model 2) that was developed by implementing feature extraction and model averaging in Analysis 2 (in the absence of a superior model group) resulted in an improved predictive performance when applied to the extrapolation test set. Although the performance of the 12-feature model on the extrapolation test set was reduced compared to that on the training set, the discrepancy could be considered to be largely due to the model having to go beyond its scope to make predictions on the test set, and less a result of model overfitting as several methods (LASSO regression, repeated ten-fold cross-validation, feature extraction and model averaging) implemented specifically to limit this.

There are several limitations to the present study. Although the fair predictive performance (AUC = 0.75) that was achieved by the prediction model when applied to the extrapolation test set was promising given that it involved going beyond the scope of the model, It would be more ideal for the model to be able to achieve a good predictive performance, which would require an AUC of at least 0.8 [340], [341]. A limitation in the study design is that models were trained on MDD patients given the limited availability of subthreshold individuals who developed MDD in the dataset. In relation to this, model reproducibility is anticipated to improve if training is to be conducted on the latter group. Small sample size is a major

limitation in many psychiatric studies, due to the general difficulty associated with recruiting appropriate patient and reference samples. To ensure sufficient sample size for the present analysis, a relatively liberal definition of subthreshold depression was employed compared to other studies, allowing for individuals experiencing mild and/or infrequent symptoms to be included as long as they fulfilled the specified criteria. Additionally, a period of up to four years between initial assessment and subsequent diagnosis of MDD was allowed, but predictive performance may improve if a shorter period is examined. Finally, although this work aimed to conduct a comprehensive analysis of the various features that could be associated with MDD outcome, other potentially important features may have been overlooked.

In conclusion, the prediction of future onset of depression in subthreshold symptomatic individuals was investigated using their proteomic, sociodemographic and clinical data. A parsimonious 12-feature prediction model was developed in the presence of model selection uncertainty by applying feature extraction and model averaging based on a set of candidate models. The results of the present study suggest that early manifestations of depression, as represented by a combination of serum proteins, sociodemographic factors and depressive symptoms, can be detected in subthreshold individuals up to four years prior to clinical diagnosis. Having demonstrated that subthreshold individuals who developed MDD could be differentiated from those who did not develop MDD, further studies need to be conducted in subthreshold individuals for a better identification and characterisation of the condition to enable earlier interventions and improved outcomes.

Chapter 5 Diagnostic prediction model development using data from dried blood spot proteomics and a digital mental health assessment to identify major depressive disorder among individuals presenting with subthreshold depression

5.1. Introduction

MDD is a highly debilitating disorder that affects more than 300 million people worldwide [2]. It is characterised by low mood and energy levels, and is associated with a reduced quality of life, increased mortality rates and high economic burdens [6], [10], [12] (see **Chapter 1**). Given the limited biological disease understanding of MDD, diagnosis currently relies on the clinical evaluation of self-reported symptoms, with no objective tests. The diagnostic criteria, including the number and duration of symptoms required, are outlined in formal classification systems, such as the DSM-5 [1] and the ICD-11 [16]. However, as clinical evaluations by GPs in primary care, in which the majority of care for depression is delivered, are usually time-restrained and can be subjective, MDD is frequently under-, over- or misdiagnosed. A meta-analysis of over 50,000 patients found that only about 47% of MDD patients are correctly identified by GPs [17]. As a result, patients are likely to receive no or inappropriate treatment, often for several years, which is detrimental to patient outcomes (see **Chapter 1**). Thus, there is a clear clinical need to develop an objective, reliable and accessible test to facilitate earlier recognition and more accurate diagnosis of MDD, and thereby enable more effective care for patients and reduce the disease burden.

As discussed in **Chapter 1**, there is a growing interest in identifying blood-based proteomic biomarkers of MDD, not only due to their ease of accessibility, but also given the converging evidence that disease-related alterations can be detected in the peripheral system, such as the HPA axis, immune/inflammatory response and metabolism [14], [222]–[227]. While clinical proteomic studies have traditionally relied on the use of serum and plasma as the sample source of blood, DBS offers a novel and innovative sampling technique with several advantages [305]. These include easy and minimally invasive sample collection (*e.g.*, self-collection by means of a finger prick), small blood volume requirement, convenient sample storage and shipment (standard postage), and lower cost [202], [303], [306] (see **Section 2.2.2**). Although yet largely unexplored, the implementation of DBS sampling for proteomic biomarker discovery and validation holds great promise for the development of a non-invasive and cost-effective diagnostic test for MDD, especially as the possibility of remote/home self-sample collection could facilitate patient recruitment which is notoriously challenging for psychiatric disorders [202], [310].

Furthermore, advances in information technology present opportunities to improve patient care (disease detection, monitoring and management) by enabling increased access to patient data [95] (see **Chapter 1**). Various patient information that are potentially related to disease status or outcome, including sociodemographic characteristics, symptom profiles and personality traits, can be readily collected through web- or smartphone-based digital platforms. The extensive patient data can be combined with biomarkers to build risk prediction models that estimate the probabilities or risks of individuals having or developing MDD [254], [255]. When used within a clinical setting, such models could complement evaluations by healthcare professionals and assist them to make more statistically-informed and objective decisions about the diagnosis of patients [256] (see **Chapter 1**).

The aim of this chapter was to develop risk prediction models with the potential to be used as a diagnostic aid for MDD, based on data from DBS proteomics and a digital mental health assessment. Individuals presenting with subclinical low mood (subthreshold depression) were recruited to investigate a clinically relevant population, the rationale behind which was analogous to that adopted in **Chapter 4**. Diagnostic prediction models of MDD were developed by comparing established current MDD patients and low mood controls, and the models were subsequently evaluated by applying them to differentiate between new current MDD patients and low mood controls, as well as between established non-current MDD patients and low

mood controls. The multimodel method of **Chapter 3** was used for prediction model development, and a repeated nested cross-validation approach was adopted to evaluate variation in model selection and ensure model reproducibility.

In the work reported in this chapter, blood samples were prepared and run on the MS by by Nitin Rustogi. Raw MS data were processed by Dr. Santiago Lago using Skyline. I was provided with the peak area values of the endogenous and SIL peptide-transitions and conducted all of the data pre-processing and analysis myself using R.

5.2. Materials and methods

5.2.1. Clinical samples

This study (International Registered Report Identifier (IRRID): RR2-10.2196/18453) investigated participants from the Delta Study, a study launched in April 2018 by the Cambridge Centre for Neuropsychiatric Research at the University of Cambridge in collaboration with Psyomics Ltd. (IRRID: RR1-10.2196/18453). The aim of the Delta Study was to develop and evaluate tests, based on data from DBS proteomics and a novel digital mental health assessment, to be used as diagnostic aids for mood disorders (BD and MDD) in individuals presenting with subclinical low mood [296]. The primary objective of the Delta Study was to reduce the misdiagnosis of BD as MDD, while the secondary objective of the Delta Study was to achieve earlier and more accurate diagnosis of MDD; the present chapter was focused on the secondary objective. The Delta Study was approved by the University of Cambridge Human Biology Research Ethics Committee (approval number HBREC 2017.11) and conducted under the standards of Good Clinical Practice and in compliance with the principles of the Declaration of Helsinki [298]. A detailed description of the Delta Study research protocol can be found in *Olmert et al. (2020)* [296].

The study flow chart is shown in **Figure 5.1**. A total of 5,422 participants were recruited online through e-mail, CCNR website and Facebook. Written informed consent was obtained from all participants upon enrolment. Inclusion criteria for the Delta Study were: age 18-45; UK resident; not pregnant or breastfeeding; not suicidal; and presenting with at least 'low mood' at the time of recruitment, as determined by the Patient Health Questionnaire-9 (PHQ-9) [373],

[374]. The PHQ-9 measured the severity of nine DSM-5 depressive symptoms in the past two weeks on a scale of zero (none) to three (severe); PHQ-9 total score ≥ five was required for inclusion in the study (depression severity classification based on total score: 0 – 4 = none; 5 – 9 = mild; 10 – 14 = moderate; 15 – 19 = moderately severe; 20 – 27 = severe).

Three thousand, two hundred and thirty two participants completed the digital mental health assessment (available through the Delta Study website), which collected data on their diagnosis history, sociodemographic and lifestyle characteristics, symptom profiles and personality traits. Participants' diagnosis history (combined with their current disease status) was used to select participants for the clinical groups, whereas other non-overlapping variables were used as predictors for model selection (as described below in **Section 5.2.3**). DBS sample collection kits were sent by post to eligible participants who consented to providing a DBS sample and completing a telephone diagnostic interview, were free from blood-borne illnesses, and had no previous diagnosis of schizophrenia. One thousand, three hundred and seventy-seven participants provided DBS samples (details on DBS sample collection are outlined below in **Section 5.2.4**).

In those who successfully completed the digital questionnaire and provided DBS samples, diagnoses of MDD and BD were determined using the WHO World Mental Health CIDI, version 3.0 [25], conducted by specially trained interviewers over the telephone. All interviewers received in-person training and thereafter were certified by an external licensed CIDI trainer to administer the CIDI. Only sections of the CIDI required for mood disorder diagnosis (screening, depression and mania sections) were implemented. Nine hundred and twenty-four participants attended and completed the telephone diagnostic interviews.

5.2.2. Study design

Overall, 897 participants completed the digital mental health assessment, provided usable DBS samples, were fasting at the time of sample collection, and obtained CIDI diagnoses (**Figure 5.1**). To be eligible for selection in the present study (to achieve the secondary objective), additional criteria to the general Delta Study inclusion and exclusion criteria were imposed. Participants had to have no previous or new diagnosis of BD, based on their self-reported diagnosis history and CIDI results, respectively. Moreover, participants had to be experiencing at least 'subthrehold depression', defined as presenting with two or more depressive symptoms,

including at least one of sadness or anhedonia (*i.e.*, two core symptoms of the DSM), whereby a symptom was considered as present if the corresponding PHQ-9 item was above zero (see **Chapter 4**).

For the purpose of this study, 295 participants were selected based on their current disease status (determined by the CIDI) and self-reported diagnosis history (collected from the digital assessment). They were: 130 subclinical low mood controls (no lifetime MDD), 40 currently depressed individuals with a new MDD diagnosis (new current MDD; MDE within the past month, no previous diagnosis of MDD, unmedicated), 53 currently depressed individuals with an existing MDD diagnosis (established current MDD; MDE within the past month, previous diagnosis of MDD, 33 on antidepressant medication), and 72 currently not depressed individuals with an existing MDD diagnosis (established non-current MDD; MDE within the past six months but not within the past month, previous diagnosis of MDD, 47 on antidepressant medication).

Developing diagnostic prediction models of MDD would ideally involve investigating differences between new current MDD patients and low mood controls. However, due to the limited number of new current MDD patients in the dataset, prediction models were first trained to differentiate between established current MDD patients and low mood controls (training set), and subsequently applied to differentiate between new current MDD patients and low mood controls (extrapolation test set), as well as between established non-current MDD patients and low mood controls (application set). The reference group of low mood controls (n = 130) was randomly split into 2:1 (matched for sex and age distribution) to form the training (n = 87) and extrapolation test/application (n = 43) sets.

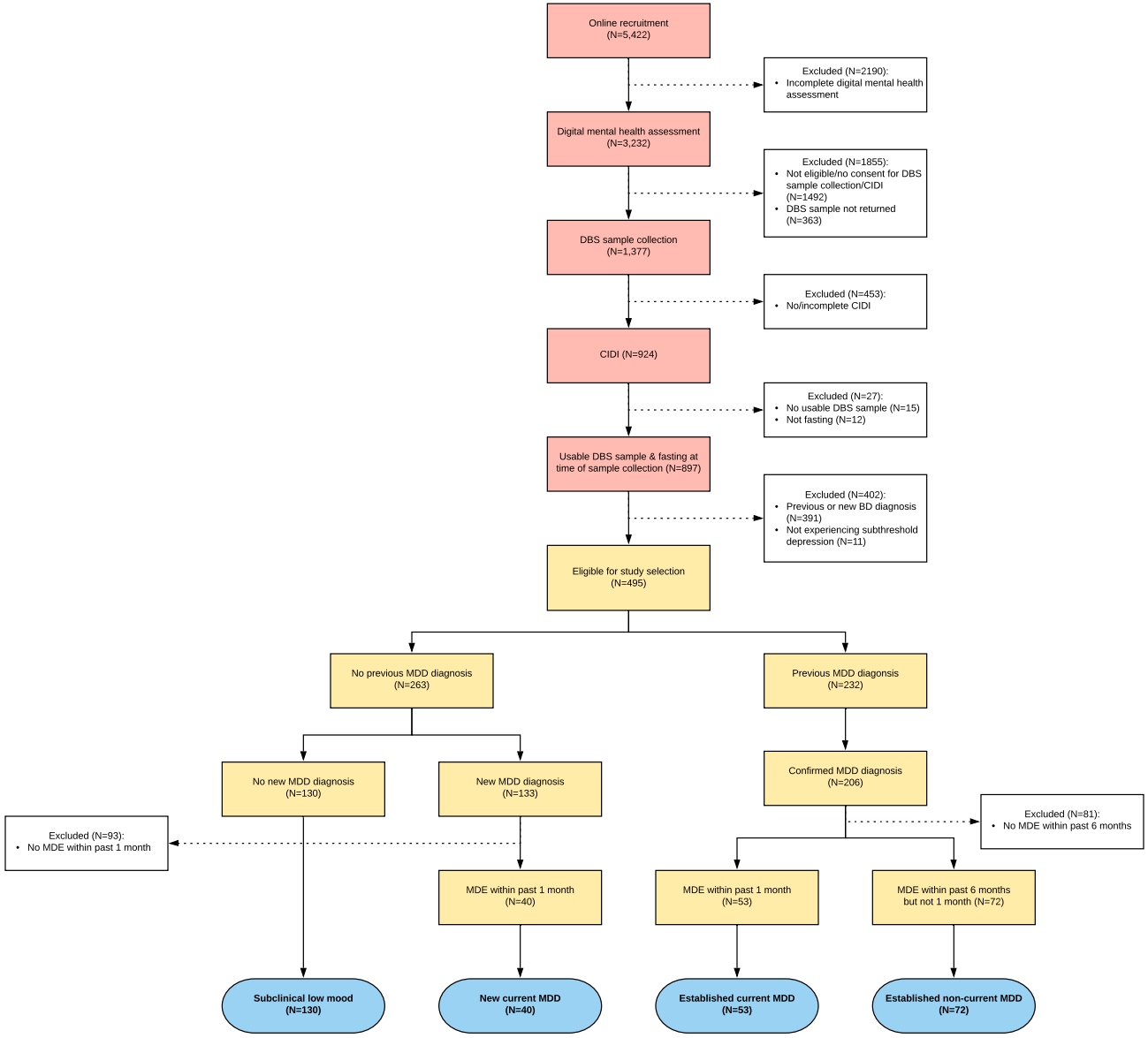

Online recruitment (N=5,422)
Excluded (N=2190):
• Incomplete digital mental health assessment
Digital mental health assessment (N=3,232)
Excluded (N=1855):
• Not eligible/no consent for DBS sample collection/CIDI (N=1492)
• DBS sample not returned (N=363)
DBS sample collection (N=1,377)
Excluded (N=453):
• No/incomplete CIDI
CIDI (N=924)
Excluded (N=27):
• No usable DBS sample (N=15)
• Not fasting (N=12)
Usable DBS sample & fasting at time of sample collection (N=897)
Excluded (N=402):
• Previous or new BD diagnosis (N=391)
• Not experiencing subthreshold depression (N=11)
Eligible for study selection (N=495)
No previous MDD diagnosis (N=263)
Previous MDD diagonsis (N=232)
No new MDD diagnosis (N=130)
New MDD diagnosis (N=133)
Confirmed MDD diagnosis (N=206)
Excluded (N=81):
• No MDE within past 6 months
Excluded (N=93):
• No MDE within past 1 month
MDE within past 1 month (N=40)
MDE within past 1 month (N=53)
MDE within past 6 months but not 1 month (N=72)
Subclinical low mood (N=130)
New current MDD (N=40)
Established current MDD (N=53)
Established non-current MDD (N=72)

Figure 5.1. Delta Study participant selection and attrition flow chart.
The number of participants in each step of the study and the reasons for attrition are shown. The steps relevant to the Delta Study are coloured in red; the steps relevant to the present study are coloured in yellow; the selected participants are coloured in blue. Abbreviations: BD (bipolar disorder); CIDI (World Health Organization World Mental Health Composite International Diagnostic Interview); DBS (dried blood spot); MDD (major depressive disorder); MDE (major depressive episode).

5.2.3. Digital mental health assessment

Data of participants' sociodemographics and lifestyle characteristics, symptom profiles and personality traits were collected from the digital mental health assessment. Sociodemographic and lifestyle factors included: sex, age, BMI, higher education (undergraduate or above), employment status, stable relationship, smoking, alcohol consumption, recreational drug use, alcohol- or drug-related problems, self-rated physical health and mental health, chronic disease, family history of psychiatric disease, childhood trauma and major life event in the past six months (**Table 5.1**). Depressive symptoms and mental well-being features were derived from the PHQ-9 [373], [374] and the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) [375], respectively. Assessment of personality traits was based on the NEO Five-Factor Inventory (NEO FFI), in which each of the five domains of personality ('Big Five'; agreeableness, conscientiousness, extraversion, neuroticism, openness) was measured using 12 items [376]. A total of 102 variables were derived from the digital mental health assessment and included as predictors for model selection.

The sociodemographic and lifestyle characteristics of participants in different clinical groups were compared using Kruskal-Wallis *H* tests for numerical variables and χ^2 tests for categorical variables (see **Section 2.5.6**). *P*-values < 0.05 were considered to be statistically significant (**Table 5.1**).

5.2.4. Targeted protein quantification

5.2.4.1. DBS sample collection

DBS sample collection kits containing relevant materials and instructions were sent to addresses specified by participants. The kit used was a Conformité Européenne marked device

under Article 22 of the Medical Device Regulation 2017/745. Participants were requested to collect samples after fasting for at least six hours. To collect DBS samples, participants cleaned a finger with an alcohol pad (after washing hands with soap and warm water); performed a finger prick using a sterile lancet; after removing the first drop of blood with a cotton ball, spotted a single blood drop per circle onto Perkin Elmer 226 Spot Saver Cards (Waltham, Massachusetts, USA) to collect five blood spot samples; and air-dried the blood spots for at least three hours at room temperature. Samples were then returned to the CCNR laboratory by standard post (sealed in a storage bag containing a desiccant) and stored with dessicants at room temperature prior to analysis.

5.2.4.2. DBS sample preparation

Two hundred and ninety-five DBS samples of the selected participants were randomised to allocate equal numbers of patient and reference groups across the experimental plates for MS analysis. Samples were prepared in a 96-well plate format as described in **Section 2.3.1.2**.

Note that overall, almost 900 DBS samples were prepared and analysed across 12 experimental plates for the Delta Study (among which 295 DBS samples were relevant to this chapter; the majority of the remaining DBS samples were relevant to the primary objective).

5.2.4.3. LC-MS/MS analysis

One hundred and ninety-four peptides representing 115 proteins, the majority previously associated with psychiatric disorders, were measured using MRM-MS. Peptides were quantified at the transition level. More details on LC-MS/MS analysis can be found in **Section 2.3.2.** The researchers conducting the sample preparation and the MS analysis were blinded to the clinical status of the participants.

5.2.4.4. QC samples

Pooled and volunteer DBS QC samples were used to assess the technical variation associated with instrument performance and sample preparation as described in **Section 2.3.3**. A pooled QC sample was prepared by pooling together the digested clinical DBS samples from the first four plates and injected once a day along with the clinical samples for the duration of the entire study (106 injections in total). Thirty-nine DBS samples obtained from a healthy volunteer were prepared following the same protocol as the clinical samples and distributed across the

experimental plates. More details on targeted protein quantification can be found in **Section 2.3**.

5.2.5. MS data pre-processing

Pre-processing of the MS data was carried out as described in **Chapter 2** (see **Section 2.4**). Peptide-transition quantification was based on the relative abundance of the endogenous and the SIL peptide-transitions, reported as the abundance ratio.

5.2.5.1. Median normalisation

As targeted protein quantification in the Delta Study involved processing and LC-MS/MS analysis of DBS samples across 12 experimental plates, the potential for plate effects was tested. Kruskal-Wallis tests by ranks (a non-parametric one-way ANOVA) were used to test for any biases arising from processing and analysing samples across multiple experimental plates (See **Section 2.5.6**). The null hypothesis was that there was no statistically significant difference in peptide-transition abundance ratios between the plates. Plate effects were considered as significant when *p*-values were < 0.05. As all peptide-transitions reported *p*-values < 0.01, median normalisation was conducted to correct for the differences between plates. This method scales the data so that there is no difference in the median values between plates [377].

For each peptide-transition, the median scaling factor per plate was estimated by scaling (dividing) the ratio abundance values of the relevant samples (*i.e.*, within the plate) by the median and then calculating the median of the scaled values. The normalised value for each peptide-transition per plate was calculated by applying the median scaling factor per plate to the relevant samples:

$$X_{norm} = \frac{X_{raw}}{median\ scaling\ factor}$$

where X_{raw} is the raw ratio abundance value and X_{norm} is the normalised ratio abundance value.

After normalisation by median scaling, all peptide-transitions reported *p*-values > 0.05; that is, there was no significant difference in the median peptide-transition abundance ratios of the clinical samples between plates. The normalised abundance ratio was then $\log_2$-transformed for statistical analysis.

5.2.5.2. Quantifier transition selection

Quantifier transitions were selected for each peptide as the transition with the highest abundance (highest peak area value) in both the endogenous and the SIL peptides, when the difference in CV between the given transition and the transition with the lowest CV was less than 5% (based on CV values of the volunteer DBS QC samples). In 192 out of 194 peptides, the difference in CV between the most abundant transition and the transition with the lowest CV was less than 5%; therefore, the most abundant transition was selected as the quantifier transition. In the remaining two peptides (CO4A_DFALLSLQVPLK, KNG1_DIPTNSPELEETLTHTITK), the most abundant transition was not selected as the quantifier transition. Subsequent analyses were conducted using data of the quantifier transitions only. Peptide quantification was based on the relative abundance of the endogenous and SIL peptide quantifier transitions, and the abundance ratio was $\log_2$-transformed for statistical analysis.

5.2.6. Statistical quality control

PCA was conducted to identify any outliers based on the $\log_2$-transformed abundance ratios of 194 peptides (**Figure 5.2**). No outlier samples were identified.

In addition, the technical variation was assessed by calculating the geometric CVs of peptide-transition abundance ratios of the pooled (variation in MS instrument performance) and the volunteer DBS (variation in sample preparation) QC samples across MS runs, as described in **Chapter 2** (see **Section 2.4.6**). The median CV values for the pooled and the volunteer DBS QC runs were 18.3% and 16.9%, respectively. The biological variation was assessed by calculating the geometric CVs of peptide-transition abundance ratios of clinical samples within each clinical sample group. The median CV values were: 27.8% for the reference group (low mood controls), 26.0% for the training set patient group (established current MDD patients), 26.2% for the extrapolation test set patient group (new current MDD patients), and 28.0% for the application set patient group (established non-current MDD patients).

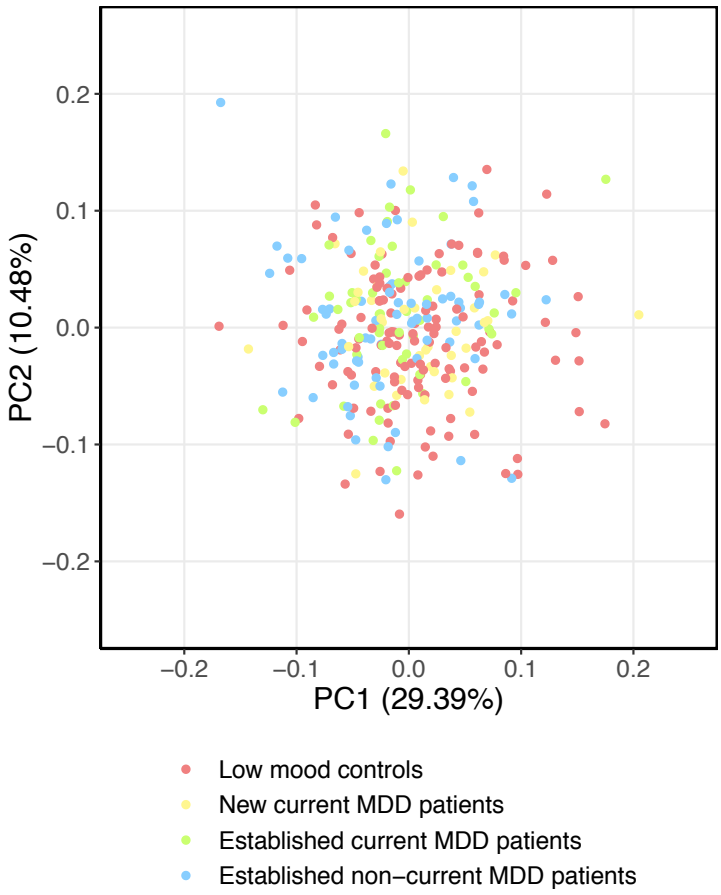


Figure 5.2. PCA plot of protein peptide abundance ratios of clinical samples.

The scores for the first two PCs are plotted with the percentage of variation accounted for by each PC shown in the axis labels. Data points are coloured according to the allocation of clinical samples into the reference group (low mood controls), the training set patient group (established current MDD patients), the extrapolation test set patient group (new current MDD patients), and the application set patient group (established non-current MDD patients). Abbreviations: MDD (major depressive disorder); PC (principal component); PCA (principal component analysis).

5.2.7. Data analysis

5.2.7.1. Model selection

A total of 296 features (194 peptides and 102 variables from the digital mental health assessment) was analysed for model selection. The full list of the analysed proteins and peptides can be found in the **Appendix** (**Table A. 1**). There were no variables with missing values. Categorical variables were represented as sets of dummy variables. Group LASSO

regression was used to allow for sets of dummy variables derived from categorical variables to be selected together. Five-fold cross-validation was used to further reduce overfitting by selecting the value of the shrinkage parameter λ that resulted in the most regularised model. More details on group LASSO regression and *K*-fold cross-validation can be found in **Chapter 2** (see **Sections 2.5.3 and 2.5.4**).

One hundred models were generated by repeatedly applying group LASSO regression with five-fold cross-validation on the training set to investigate model selection uncertainty. The Akaike weight of each model was calculated using the AICc and interpreted as the probability that the model was the best approximating model for the data. The selection fraction of each feature was measured as the proportion of models out of 100 in which it was selected and used to assess the relative importance of the features. Models which comprised of the same combinations of features selected were grouped together, and the frequency of occurrence and the summed Akaike weight of each model group were measured to estimate the probability that the selected combination of features comprised the best approximating model. More details on Akaike model weights and relative feature importance can be found in **Chapter 3** (see **Section 3.2.5**).

5.2.7.2. Feature extraction and model averaging

Due to the uncertainty in model selection, feature extraction and model averaging were implemented using all 100 models to obtain more reproducible predictions of the probability of MDD outcome. Based on the findings in **Chapter 3**, only features with selection fractions ≥ 0.9 were included in the prediction model (feature extraction) to limit overfitting, and subsequently averaged over the 100 models (model averaging) to obtain better estimates of feature coefficients [284], [289], [290]. The weighted average coefficient of a given feature was estimated across the 100 models as described in **Chapter 3** (see **Section 3.2.5.7**).

5.2.7.3. Predictive performance

Predictive performance of the models when applied to the training and test sets was evaluated by plotting ROC curves and measuring the AUC. More details on model predictive performance can be found in **Chapter 2** (see **Section 2.5.5)**.

5.2.7.4. Repeated nested cross-validation

As different random splits of the dataset (here, the reference group of low mood controls) into training and test sets can lead to variation in feature selection and model predictive performance, a repeated nested cross-validation (also known as double cross-validation) approach was adopted [260], [378] to assess the variation and obtain an unbiased and generalised estimate of model performance (**Figure 5.3**).

In nested cross-validation, the full reference group dataset (n = 130) was randomly split into three folds (matched for sex and age distribution). Each fold was retained as a test set reference group and the remaining two folds were used as a training set reference group. In the inner loop of nested cross-validation, the training set, consisting of 87 low mood controls and 53 established current MDD patients, was used to select a prediction model (based on repeated group LASSO with five-fold cross-validation, followed by feature extraction and model averaging). In the outer loop, the predictive performance of the selected model when applied to the extrapolation test set, consisting of 43 low mood controls and 40 new current MDD patients, and the application set, consisting of the same 43 low mood controls and 72 established non-current MDD patients, was assessed. Note that, while the use of low mood controls followed a nested cross-validation approach (as those used for model assessment were alternately also used for model selection), patients used for model assessment in the outer loop were never used for model selection in the inner loop. Therefore, this should be considered as an adapted nested cross-validation approach.

In repeated nested cross-validation, this procedure was repeated five times in an addiitonal repetition loop, each with a different random split of the full reference group dataset into training and test sets. This resulted in 15 prediction models (five repeats of three-fold nested cross-validation). Repetition is a key component of reliable model assessment based on nested cross-validation, as given the variability, model predictive performance cannot be reliably assessed using any single nested cross-validation run [260]. The overall importance of the features was assessed based on their frequency of appearance across the 15 models, whereby those appearing in at least five models were considered as the most robust features. A generalised estimate of model performance was obtained by finding the average performance across the 15 models (reported as the mean AUC ± 95% confidence intervals). The importance of repeated nested cross-validation in model evaluation has previously been demonstrated in *Filzmoser et al. (2009)* [378] and *Krstajic et al. (2014)* [260].

5.2.7.5. Antidepressant medication

Mann-Whitney *U* tests [343] and bootstrap hypothesis tests [379] were performed to assess whether any of the proteomic features or BMI were significantly associated with participants' current antidepressant use. The abundances of 194 protein peptides and BMI were compared between 33 antidepressant users and 20 non-users among established current MDD patients. The null hypothesis was that there was no statistically significant effect of antidepressant medication use on the value/abundance of the feature of interest. *P*-values < 0.05 were considered to be statistically significant. Additionally, unsupervised PCA was conducted to assess the effect of antidepressant use on the patients' proteomic profiles.

5.2.7.6. Associations of selected proteomic biomarkers with lifestyle factors

As lifestyle factors such as smoking, alcohol, drug use and poor self-rated physical health could affect the abundances of protein peptides and thus have potential confounding effects on the results of model selection, Mann-Whitney *U* tests [343] were performed to assess the associations of the selected proteomic biomarkers with the lifestyle factors. The protein peptide abundances of the features appearing the 15 prediction models were associated with binary measures of smoking, alcohol consumption, recreational drug use and self-rated poor physical health (yes/no) in 295 participants. The null hypothesis was that there was no statistically significant effect of the lifestyle factor on the abundance of the feature of interest. The Benjamini-Hochberg method was used to adjust for multiple testing [380]. *P*-values < 0.05 were considered to be statistically significant.

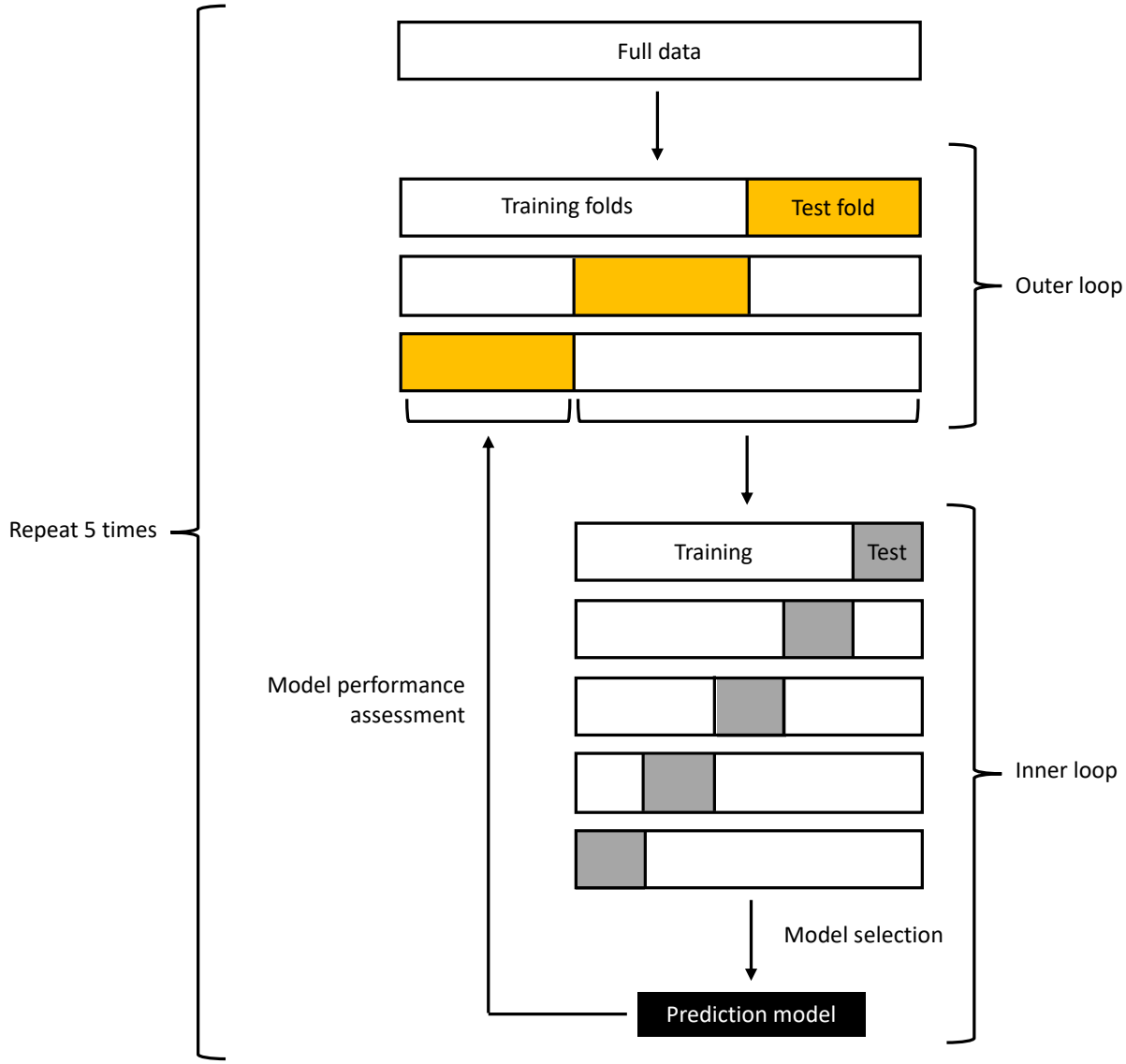


Figure 5.3. Schematic of repeated nested cross-validation.

In nested cross-validation, the full reference group dataset was randomly split into three folds. Each fold was retained as a test set reference group and the remaining two folds were used as a training set reference group. In the inner loop of nested cross-validation, the training set was used to select a prediction model. In the outer loop, the predictive performance of the selected model when applied to the extrapolation test set and the application set was assessed. In repeated nested cross-validation, this procedure was repeated five times, each with a different random split of the full reference group dataset into training and test sets, resulting in 15 models (five repeats of three-fold nested cross-validation).

5.3. Results

One hundred and ninety-four DBS proteomic and 102 digital mental health assessment features (296 in total) were measured in the reference group of 130 low mood controls, the training set patient group of 53 established current MDD patients, the extrapolation test set patient group of 40 new current MDD patients, and the application set patient group of 72 established non-current MDD patients. **Table 5.1** shows their sociodemographic and lifestyle characteristics. The participants consisted of 70.8% females, which reflects the gender difference in the prevalence of MDD (women have a two-fold increased risk of MDD than men)[30], and were on average 25.9 (standard deviation = 6.3) years old.

Repeated nested cross-validation revealed a degree of variation in feature selection across the 15 average prediction models (**Table 5.2**). Note that for a particular feature to appear in an average prediction model, it had to be selected in at least 90 out of 100 models (selection fraction ≥ 0.9) in the inner loop of nested cross-validation. The number of features in a model ranged between five and 15, with a median of nine. None of the models comprised an identical combination of features. Overall, there were 29 unique features (16 proteomic biomarkers and 13 digital features) which appeared in at least one model; the weighted average coefficients of these features are shown in **Figure 5.4**. Features that appeared in more than one model showed consistent directions of change across different models. Based on their frequency of appearance across the 15 models, the most robust features (and the number of models in which they appeared) were: A1AG1_SDVVYTDWK (8); A2GL_VAAGAFQGLR (6); AL1A1_ILDLIESGK (5); APOE_LGPLVEQGR (15); CFAH_NGFYPATR (14); BMI (10); poor mental health (15); positive emotions (extraversion) (12); tender-mindedness (agreeableness) (6). MDD was found to be associated with increased levels of α-1-acid glycoprotein 1 (A1AG1), leucine-rich α-2-glycoprotein (A2GL), apolipoprotein E (APOE) and complement factor H (CFAH), and decreased levels of retinal dehydrogenase 1 (AL1A1); as well as with poorer self-rated mental health, higher BMI, reduced daily experiences of positive emotions such as happiness and joy, and greater tender-mindedness (empathy toward others). Moreover, unsupervised PCA of the 29 features showed a moderate separation between low mood controls and MDD patients (**Figure 5.5**).

The prediction models showed relatively stable and consistent predictive performance (**Table 5.2**). On average, the models demonstrated an excellent predictive performance in differentiating between established current MDD patients and low mood controls in the training set (AUC = 0.94 ± 0.01; **Figure 5.6a**). To evaluate the ability of the models to identify currently depressed patients among individuals presenting with low mood, they were applied to differentiate between new current MDD patients and low mood controls in the extrapolation test set. Importantly, this resulted in a good predictive performance (AUC = 0.80 ± 0.01; **Figure 5.6b**). In addition, to further evaluate the diagnostic utility of the prediction models, they were applied to differentiate between established non-current MDD patients and low mood controls in the application set. The predictive performance was fair, but nearing the AUC threshold of good (AUC = 0.79 ± 0.01; **Figure 5.6c**).

In order to assess the relative contributions of the selected digital features and proteomic biomarkers in predicting MDD, prediction models consisting only of the digital features and those consisting only of the proteomic biomarkers were applied to the data sets. Digital-only models demonstrated a good predictive performance on the training set (AUC = 0.88 ± 0.01), and a fair predictive performance on the extrapolation test set (AUC = 0.79 ± 0.01) and the application set (AUC = 0.76 ± 0.02). On the other hand, biomarker-only models demonstrated a good predictive performance on the training set (AUC = 0.76 ± 0.02), but the predictive performance on the extrapolation test set (AUC = 0.56 ± 0.03) and the application set (AUC = 0.63 ± 0.03) was poor. Therefore, while the digital-only models clearly outperformed the biomarker-only models in predicting MDD, combining the digital features and the biomarkers resulted in prediction models with improved predictive performance on all data sets, although this improvement was minimal on the extrapolation test set.

Results of Mann-Whitney *U* tests and bootstrap hypothesis tests showed that none of the selected proteomic biomarkers or BMI were significantly associated with antidepressant medication use (**Table 5.3**), which suggests that these features were likely to be related to depression itself rather than to medication. In addition, PCA of protein peptide abundance ratios of established current MDD patients (training set patient group) showed minimal separation between antidepressant users and non-users (**Figure 5.7**). Moreover, results of Mann-Whitney *U* tests showed that, although there were some significant associations between the selected biomarkers and lifestyle factors of smoking, recreational drug use and poor self-rated physical health, most of these were no longer significant after adjustment for multiple

testing; only the associations of protein peptides A1AG1_SDVVYTDWK, CFAH_IDVHLVPDR and CFAH_NGFYPATR with poor physical health remained significant after adjustment for multiple testing (adjusted *p*-values < 0.05; **Table 5.4**). This raises the possibility that the identification of A1AG1 and CFAH as biomarkers of MDD could have been influenced by their associations with poor physical health rather than with the disease itself. However, the fact that poor physical health was never selected in the prediction models despite being included as a predictor for model selection suggests that the factor itself was not significantly associated with the outcome of MDD. Hence, its potential confounding effect on the selection of the biomarkers in the prediction models was considered to be insignificant.

Table 5.1. Summary of sociodemographic and lifestyle characteristics of participants in the reference group (low mood controls), the training set patient group (established current MDD patients), the extrapolation test set patient group (new current MDD patients), and the application set patient group (established non-current MDD patients).

Numerical variables are shown as the mean (standard deviation), and binary and categorical variables are shown as the percentage of participants in each category. *P*-values were from Kruskal-Wallis *H* tests for numerical variables and χ^2 tests for categorical variables. Abbreviations: BMI (body mass index); MDD (major depressive disorder); PHQ (Patient Health Questionnaire); WEMWBS (Warwick-Edinburgh Mental Well-Being Scale).

	Low mood controls (reference group)	**Established current MDD patients (training set patient group)**	**New current MDD patients (extrapolation test set patient group)**	**Established non-current MDD patients (application set patient group)**	***p*-value**
n	130	53	40	72	
Sex % (male/female)	34.6/65.4	17.0/83.0	40.0/60.0	22.2/77.8	0.022
Age (years)	25.7 (6.5)	26.2 (5.9)	23.9 (5.7)	27.2 (6.3)	0.02
BMI (kg/m^2)	24.6 (4.8)	28.4 (6.9)	26.5 (6.2)	28.1 (6.3)	< 0.001
Higher education % (yes/no)	62.3/37.7	64.2/35.8	70.0/30.0	66.7/33.3	0.814
Employment % (employed/unemployed/ student)	56.2/3.8/40.0	49.1/13.2/37.7	55/7.5/37.5	56.9/18.1/25	0.026
Stable relationship % (yes/no)	57.7/42.3	58.5/41.5	40.0/60.0	55.6/44.4	0.232
Living alone % (yes/no)	6.9/93.1	15.1/84.9	7.5/92.5	8.3/91.7	0.350
Smoking % (yes/no)	26.9/73.1	20.8/79.2	22.5/77.5	27.8/72.2	0.764
Alcohol % (yes/no)	78.5/21.5	69.8/30.2	85.0/15.0	55.6/44.4	0.001
Recreational drug use % (yes/no)	14.6/85.4	18.9/81.1	15.0/85.0	16.7/83.3	0.905

Alcohol- or drug-related problems % (yes/no)	16.9/83.1	24.5/75.5	20/80	31.9/68.1	0.098
Poor physical health % (yes/no)	16.9/83.1	32.1/67.9	10.0/90.0	30.6/69.4	0.009
Poor mental health % (yes/no)	30.8/69.2	94.3/5.7	85.0/15.0	76.4/23.6	< 0.001
Chronic disease % (yes/no)	8.5/91.5	5.7/94.3	10.0/90.0	12.5/87.5	0.603
Chronic pain % (yes/no)	21.5/78.5	20.8/79.2	10.0/90.0	26.4/73.6	0.241
Family history of psychiatric disease % (yes/no)	66.2/33.8	86.8/13.2	67.5/32.5	83.3/16.7	0.005
Childhood trauma % (yes/no)	35.4/64.6	62.3/37.7	52.5/47.5	55.6/44.4	0.002
Major life event in the past six months % (yes/no)	47.7/52.3	64.2/35.8	65.0/35.0	51.4/48.6	0.093

Table 5.2. Summary of the 15 average prediction models obtained from five times-repeated three-fold nested cross-validation.

For each average prediction model, the number of features and its performance on the training set, extrapolation test set and application set, as measured by the AUC, are shown. Proteomic features are represented in a protein_peptide format. Personality traits are represented in a factor_trait format, the factor being one of agreeableness (A), conscientiousness (C), extraversion (E), neuroticism (N), openness (O). Abbreviations: A1AG1 (α-1-acid glycoprotein 1); A2GL (leucine-rich α-2-glycoprotein); AL1A1 (retinal dehydrogenase 1); APOA4 (apolipoprotein A-IV); APOE (apolipoprotein E); AUC (area under the receiver operating characteristic curve); BMI (body mass index); CV (cross-validation); CFAH (complement factor H); CO4A (complement C4-A); CO9 (complement component C9); IF4B (eukaryotic translation initiation factor 4B); IGHG1 (immunoglobin heavy constant γ 1); IGHG3 (Immunoglobin heavy constant γ 3); KAD1 (adenylate kinase isoenzyme 1); PARK7 (protein/nucleic acid deglycase DJ-1); PSA6 (proteasome subunit α type-6).

Repeated nested CV	Average model	Number of features	Training AUC	Extrapolation test AUC	Application AUC
Repeat 1; fold 1	APOE_LGPLVEQGR + CFAH_NGFYPATR + IGHG3_DTLMISR + BMI + Poor mental health + E_positive emotions	6	0.93	0.78	0.77
Repeat 1; fold 2	A1AG1_SDVVYTDWK + A2GL_VAAGAFQGLR + AL1A1_ILDLIESGK + APOA4_LAPLAEDVR + APOE_LGPLVEQGR + CFAH_NGFYPATR + IGHG1_FNWYVDGVEVHNAK + PARK7_DGLILTSR + BMI + Poor mental health + E_positive emotions + A_trust + A_tender-mindedness + O_ideas + Total neuroticism	15	0.95	0.80	0.80
Repeat 1; fold 3	AL1A1_ILDLIESGK + APOE_LGPLVEQGR + CFAH_NGFYPATR + CO9_LSPIYNLVPVK + Poor mental health + E_positive emotions + A_tender-mindedness	7	0.92	0.79	0.77

Repeat 2; fold 1	APOE_LGPLVEQGR + CFAH_NGFYPATR + BMI + Poor mental health + E_positive emotions	5	0.93	0.78	0.76
Repeat 2; fold 2	APOE_LGPLVEQGR + CFAH_NGFYPATR + CO9_LSPIYNLVPVK + Poor mental health + E_positive emotions	5	0.91	0.80	0.79
Repeat 2; fold 3	A1AG1_SDVVYTDWK + A2GL_VAAGAFQGLR + APOE_LGPLVEQGR + CFAH_NGFYPATR + IF4B_SILPTAPR + IGHG1_FNWYVDGVEVHNAK + BMI + Poor mental health + A_straightforwardness + A_compliance + O_ideas + O_actions	12	0.95	0.82	0.83
Repeat 3; fold 1	A1AG1_SDVVYTDWK + A2GL_VAAGAFQGLR + AL1A1_ILDLIESGK + APOE_LGPLVEQGR + CFAH_NGFYPATR + IF4B_SILPTAPR + BMI + Poor mental health + E_positive emotions + A_tender-mindedness + O_ideas	11	0.95	0.81	0.83
Repeat 3; fold 2	A1AG1_SDVVYTDWK + APOE_LGPLVEQGR + CFAH_NGFYPATR + CO4A_DFALLSLQVPLK + CO9_LSPIYNLVPVK + IGHG1_FNWYVDGVEVHNAK + Poor mental health + E_positive emotions + N_depression	9	0.93	0.80	0.78
Repeat 3; fold 3	APOE_LGPLVEQGR + CFAH_NGFYPATR + PARK7_DGLILTSR + BMI + Poor mental health + A_tender-mindedness + C_competence + O_actions + Total agreeableness	9	0.92	0.80	0.78
Repeat 4; fold 1	A1AG1_SDVVYTDWK + AL1A1_ILDLIESGK + APOE_LGPLVEQGR + CFAH_IDVHLVPDR + PARK7_DGLILTSR + PSA6_HITIFSPEGR + BMI + Poor mental health + E_positive emotions + Total neuroticism	10	0.94	0.79	0.77

Repeat 4; fold 2	APOE_LGPLVEQGR + CFAH_NGFYPATR + BMI Poor mental health + E_positive emotions + A_tender-mindedness	6	0.92	0.82	0.80
Repeat 4; fold 3	A1AG1_SDVVYTDWK + A2GL_VAAGAFQGLR + APOE_LGPLVEQGR + CFAH_NGFYPATR + IF4B_SILPTAPR + IGHG3_DTLMISR + KAD1_IIFVVGGPGSGK + Poor mental health + E_positive emotions + O_ideas + O_actions	11	0.95	0.82	0.81
Repeat 5; fold 1	APOA4_ISASAEELR + APOE_LGPLVEQGR + CFAH_NGFYPATR + IGHG3_DTLMISR + BMI + Poor mental health + E_positive emotions	7	0.94	0.76	0.73
Repeat 5; fold 2	A1AG1_SDVVYTDWK + A2GL_VAAGAFQGLR + APOE_LGPLVEQGR + CFAH_NGFYPATR + BMI + Poor mental health + Total neuroticism	7	0.92	0.82	0.80
Repeat 5; fold 3	A1AG1_SDVVYTDWK + A2GL_VAAGAFQGLR + AL1A1_ILDLIESGK + APOE_LGPLVEQGR + CFAH_NGFYPATR + CO9_LSPIYNLVPVK + IGHG1_FNWYVDGVEVHNAK + Poor mental health + E_positive emotions + A_tender-mindedness	10	0.95	0.80	0.79

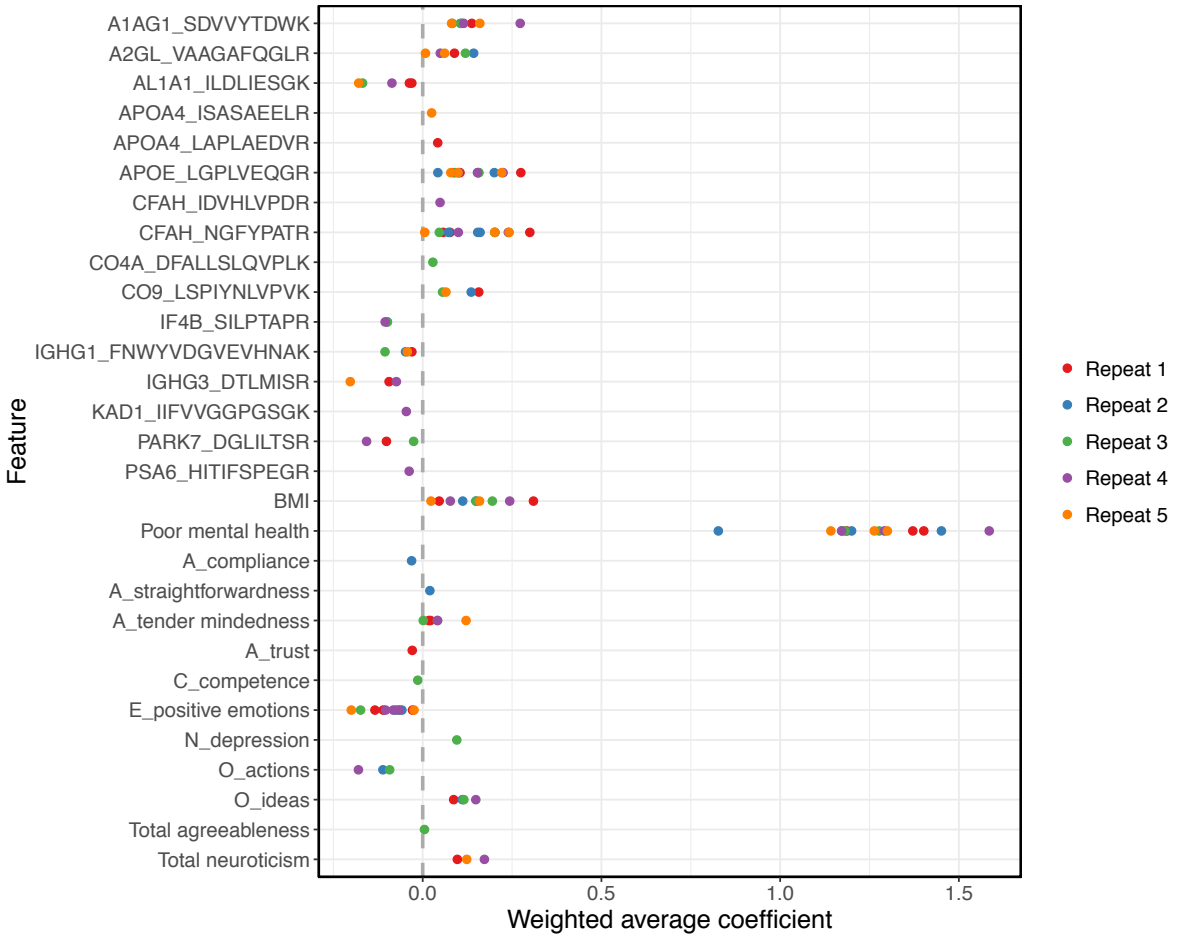


Figure 5.4. Coefficients of features appearing in the 15 average prediction models.

Features with coefficients > 0 were increased in MDD patients compared to low mood controls, and features with coefficients < 0 were decreased in MDD patients compared to low mood controls. The number of datapoints of each feature represents the number of prediction models in which it appears. Data points are coloured by the repeat number allocated to the models in repeated nested cross-validation. Proteomic features are represented in a protein_peptide format. Personality traits are represented in a factor_trait format, the factor being one of agreeableness (A), conscientiousness (C), extraversion (E), neuroticism (N), openness (O). Abbreviations: A1AG1 (α-1-acid glycoprotein 1); A2GL (leucine-rich α-2-glycoprotein); AL1A1 (retinal dehydrogenase 1); APOA4 (apolipoprotein A-IV); APOE (apolipoprotein E); BMI (body mass index); CFAH (complement factor H); CO4A (complement C4-A); CO9 (complement component C9); IF4B (eukaryotic translation initiation factor 4B); IGHG1 (immunoglobin heavy constant γ 1); IGHG3 (Immunoglobin heavy constant γ 3); KAD1 (adenylate kinase isoenzyme 1); MDD (major depressive disorder); PARK7 (protein/nucleic acid deglycase DJ-1); PSA6 (proteasome subunit α type-6).

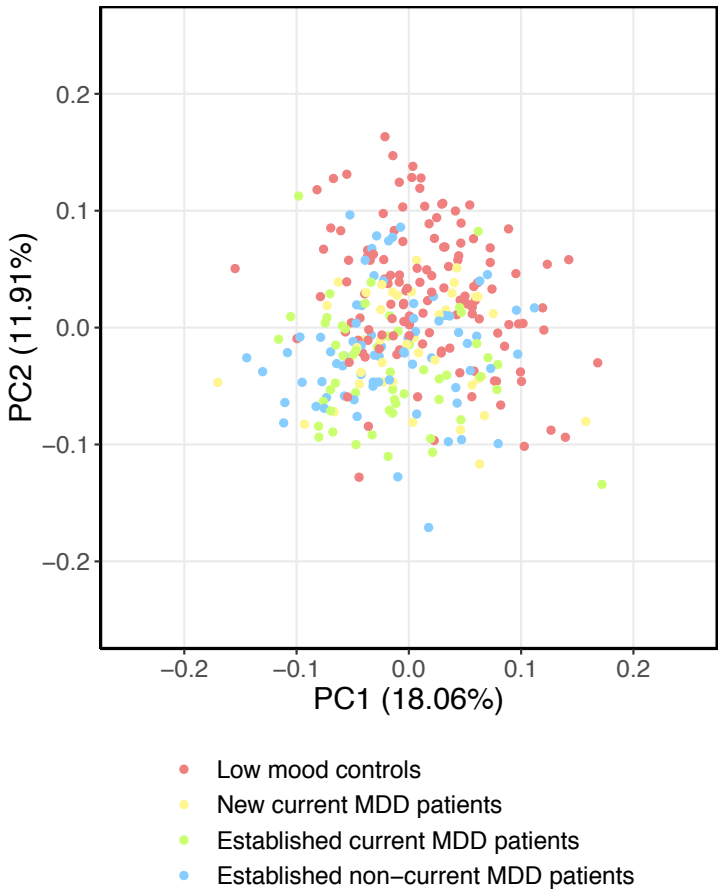


Figure 5.5. PCA plot of 29 features appearing in the 15 prediction models.

The scores for the first two PCs are plotted with the percentage of variation accounted for by each PC shown in the axis labels. Data points are coloured according to the allocation of clinical samples into the reference group (low mood controls), the training set patient group (established current MDD patients), the extrapolation test set patient group (new current MDD patients), and the application set patient group (established non-current MDD patients). Abbreviations: MDD (major depressive disorder); PC (principal component); PCA (principal component analysis).

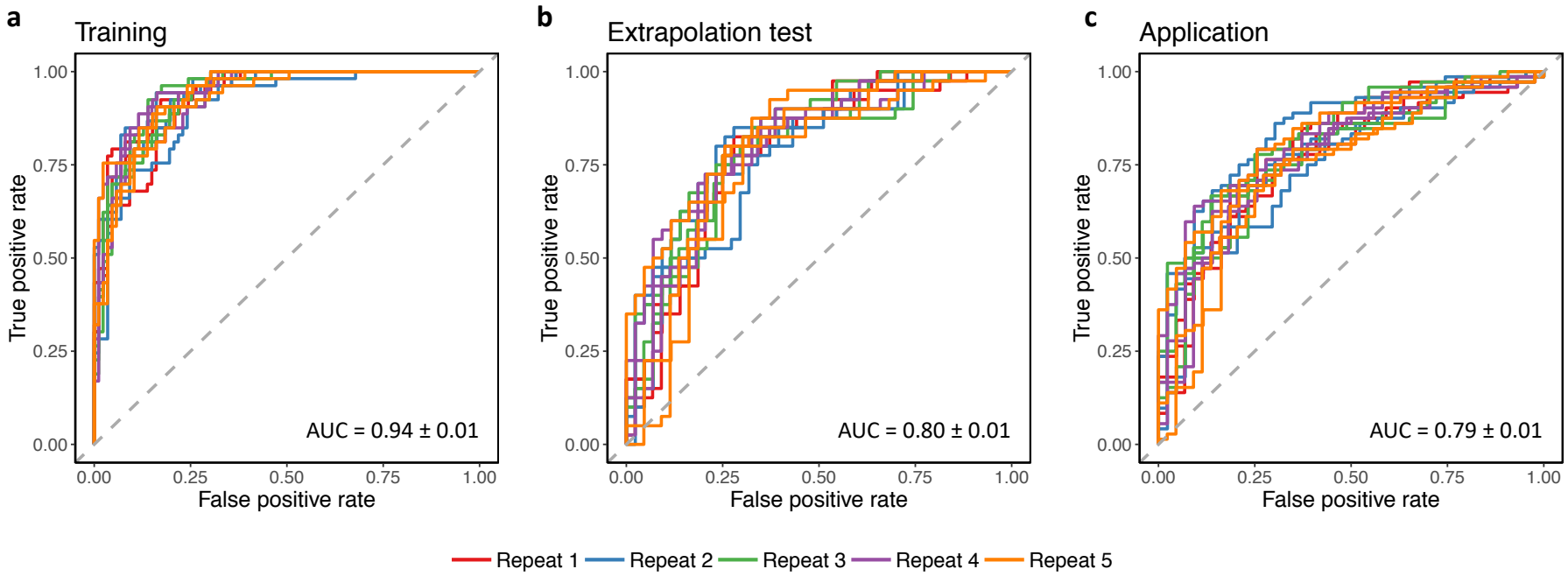


Figure 5.6. ROC curves showing model performance in predicting the probability of MDD outcome.

The prediction models were applied to predict the probability of MDD in: (a) the training set (53 established current MDD patients *vs* 87 low mood controls), (b) the extrapolation test set (40 new current MDD patients *vs* 43 low mood controls), and (c) the application set (72 established non-current MDD patients *vs* 43 low mood controls). The AUC values are shown as the mean ± 95% confidence intervals of the 15 prediction models (**Table 5.2**). ROC curves are coloured by the repeat number allocated to the models in repeated nested cross-validation. Abbreviations: AUC (area under the curve); MDD (major depressive disorder); ROC (receiver operating characteristic).

Table 5.3. Mann-Whitney *U* test and bootstrap hypothesis test results to assess the associations of protein peptide abundances and BMI with antidepresant use.

The abundances of 194 protein peptides and BMI were compared between 33 antidepressant useres and 20 non-users among established current MDD patients (training set patient group). The null hypothesis was that there was no statistically significant effect of antidepressant medication use on the value/abundance of the feature of interest. The *W*-value (Mann-Whitney *U* statistic), *p*-value and *q*-value (FDR-adjusted *p*-value) are shown for each Mann-Whitney *U* test, and the *t*-value (boostrap test statistic), *p*-value and *q*-value are shown for each bootstrap hypothesis test. Features appearing in the 15 prediction models are shown in bold, none of which had *p*-values < 0.05. Proteomic features are represented in a protein_peptide format, and their abbreviations can be found in the **Appendix** (**Table A. 1**). Abbreviations: BMI (body mass index); FDR (false discovery rate); MDD (major depressive disorder).

Feature	Mann-Whitney *U* test			Bootstrap hypothesis test		
	W-value	*p*-value	*q*-value	*t*-value	*p*-value	*q*-value
A1AG1_SDVVYTDWK	**335**	**0.9342**	**0.9947**	**-0.239**	**0.591**	**0.8446**
A1AT_LSITGTYDLK	342	0.8329	0.9718	0.6265	0.241	0.8166
A1AT_SPLFMGK	393	0.2514	0.9718	1.306	0.092	0.7176
A1AT_SVLGQLGITK	387	0.2999	0.9718	1.1589	0.115	0.767
A1BG_ATWSGAVLAGR	399	0.2088	0.9718	1.3705	0.071	0.7176
A1BG_GVTFLLR	354	0.6663	0.9718	0.4504	0.297	0.8166
A1BG_LLELTGPK	364	0.5387	0.9718	0.3616	0.324	0.8166
A1BG_SGLSTGWTQLSK	369	0.4799	0.9718	0.7063	0.215	0.8166
A2AP_DFLQSLK	314	0.7761	0.9718	-0.1153	0.557	0.842
A2AP_FDPSLTQR	348	0.7481	0.9718	0.1655	0.424	0.8166
A2AP_LFGPDLK	297	0.5509	0.9718	-0.5736	0.712	0.8648
A2GL_VAAGAFQGLR	**254**	**0.1659**	**0.9718**	**-0.6114**	**0.68**	**0.8446**
A2MG_AIGYLNTGYQR	343	0.8186	0.9718	0.582	0.273	0.8166
A2MG_NEDSLVFVQTDK	334	0.9488	0.9947	0.0246	0.466	0.8166
AACT_ADLSGITGAR	342	0.8329	0.9718	0.4364	0.312	0.8166
AACT_EIGELYLPK	331	0.9927	0.9978	0.3566	0.339	0.8166
AACT_EQLSLLDR	301	0.601	0.9718	0.1676	0.405	0.8166
AL1A1_ILDLIESGK	**280**	**0.3637**	**0.9718**	**-0.8701**	**0.81**	**0.913**
ALBU_AAFTECCQAADK	469	0.011	0.9718	2.703	0.002	0.39
ALBU_ETYGEMADCCAK	463	0.015	0.9718	2.5382	0.004	0.39

ALBU_QNCELFEQLGEYK	397	0.2224	0.9718	1.3117	0.078	0.7176
ALBU_YLYEIAR	352	0.6932	0.9718	0.6351	0.249	0.8166
ALDOA_ALANSLACQGK	388	0.2914	0.9718	1.0722	0.13	0.78
ALDOA_QLLLTADDR	317	0.8186	0.9718	-0.0161	0.517	0.8209
AMBP_ETLLQDFR	291	0.4799	0.9718	-1.2416	0.901	0.9377
AMBP_TVAACNLPIVR	389	0.2831	0.9718	0.8624	0.213	0.8166
ANGT_ALQDQLVLVAAK	331	0.9927	0.9978	-0.0959	0.54	0.8357
ANGT_SLDFTELDVAAEK	299	0.5757	0.9718	-0.2604	0.599	0.8446
ANT3_EVPLNTIIFMGR	416	0.1167	0.9718	1.9699	0.023	0.7176
ANT3_FATTFYQHLADSK	376	0.4038	0.9718	0.8952	0.18	0.8166
ANT3_FDTISEK	397	0.2224	0.9718	1.4383	0.078	0.7176
ANT3_LPGIVAEGR	390	0.2749	0.9718	1.1085	0.144	0.78
APOA1_ATEHLSTLSEK	385	0.3173	0.9718	1.306	0.084	0.7176
APOA1_LLDNWDSVTSTFSK	371	0.4574	0.9718	0.7343	0.233	0.8166
APOA1_VSFLSALEEYTK	342	0.8329	0.9718	0.385	0.317	0.8166
APOA2_EQLTPLIK	346	0.7761	0.9718	0.1544	0.418	0.8166
APOA2_SPELQAEAK	407	0.1604	0.9718	1.6085	0.052	0.7176
APOA4_ALVQQMEQLR	258.5	0.1926	0.9718	-1.2087	0.871	0.9276
APOA4_IDQNVEELK	270	0.2749	0.9718	-1.1025	0.842	0.9224
APOA4_ISASAEELR	**270**	**0.2749**	**0.9718**	**-1.1139**	**0.852**	**0.923**
APOA4_LAPLAEDVR	**263**	**0.2224**	**0.9718**	**-1.2125**	**0.876**	**0.9276**
APOA4_LLPHANEVSQK	265	0.2366	0.9718	-1.1426	0.868	0.9276
APOB_TGISPLALIK	230	0.0679	0.9718	-2.1149	0.988	0.988
APOC1_EFGNTLEDK	329	0.9927	0.9978	0.0097	0.466	0.8166
APOC1_EWFSETFQK	353	0.6797	0.9718	0.0116	0.483	0.819
APOC2_TAAQNLYEK	320	0.8616	0.9756	0.3181	0.355	0.8166
APOC2_TYLPAVDEK	292	0.4914	0.9718	-0.0902	0.494	0.8209
APOC3_GWVTDGFSSLK	342	0.8329	0.9718	0.3497	0.354	0.8166
APOD_NILTSNNIDVK	353	0.6797	0.9718	0.2253	0.401	0.8166
APOD_VLNQELR	396	0.2294	0.9718	1.4207	0.067	0.7176
APOE_AATVGSLAGQPLQER	392	0.2591	0.9718	1.094	0.141	0.78
APOE_ALMDETMK	367	0.503	0.9718	0.7367	0.221	0.8166
APOE_LGPLVEQGR	**377**	**0.3935**	**0.9718**	**1.1747**	**0.113**	**0.767**
APOE_SELEEQLTPVAEETR	369	0.4799	0.9718	0.6519	0.26	0.8166
APOH_EHSSLAFWK	306	0.6663	0.9718	-1.1701	0.885	0.9278
APOL1_LNILNNNYK	349	0.7343	0.9718	0.4844	0.324	0.8166

APOL1_VNEPSILEMSR	342	0.8329	0.9718	0.2699	0.409	0.8166
APOM_FLLYNR	348	0.7481	0.9718	0.6653	0.211	0.8166
APOM_SLTSCLDSK	381	0.3541	0.9718	1.0908	0.118	0.767
C1QB_GNLCVNLMR	365	0.5267	0.9718	0.757	0.229	0.8166
C1QC_FQSVFTVTR	346	0.7761	0.9718	0.0572	0.469	0.8166
C1QC_TNQVNSGGVLLR	351	0.7068	0.9718	0.4469	0.316	0.8166
C1R_YTTEIIK	319	0.8472	0.9718	-0.3141	0.604	0.8446
C1RL_VVVHPDYR	401	0.1958	0.9718	1.5515	0.065	0.7176
C1S_LLEVPEGR	296	0.5387	0.9718	-0.6442	0.74	0.8799
C4BPA_EDVYVVGTVLR	309	0.7068	0.9718	-0.58	0.711	0.8648
C4BPA_GYILVGQAK	369	0.4799	0.9718	0.5856	0.262	0.8166
C4BPA_YTCLPGYVR	429	0.0707	0.9718	1.779	0.051	0.7176
CAH1_ADGLAVIGVLMK	276	0.3263	0.9718	-0.6993	0.756	0.8881
CAH2_GGPLDGTYR	338	0.8905	0.9756	0.1008	0.44	0.8166
CAH2_SADFTNFDPR	355	0.653	0.9718	0.1351	0.432	0.8166
CAH2_VVDVLDSIK	353	0.6797	0.9718	0.1952	0.416	0.8166
CATA_LNVITVGPR	346	0.7761	0.9718	0.5931	0.288	0.8166
CBG_GTWTQPFDLASTR	294	0.5148	0.9718	-0.6271	0.719	0.8655
CD44_YGFIEGHVVIPR	375	0.4142	0.9718	0.5201	0.316	0.8166
CD5L_IWLDNVR	386	0.3085	0.9718	1.2269	0.09	0.7176
CD5L_LVGGLHR	428	0.0736	0.9718	1.8931	0.027	0.7176
CERU_DIASGLIGPLIICK	392	0.2591	0.9718	0.829	0.199	0.8166
CERU_EYTDASFTNR	313	0.7621	0.9718	-0.2189	0.574	0.8446
CFAB_DISEVVTPR	256.5	0.1804	0.9718	-1.2925	0.904	0.9377
CFAB_DLLYIGK	290	0.4686	0.9718	-0.8949	0.834	0.9224
CFAB_EELLPAQDIK	285	0.4142	0.9718	-0.8476	0.805	0.913
CFAB_YGLVTYATYPK	311	0.7343	0.9718	-0.2355	0.593	0.8446
CFAH_IDVHLVPDR	**289**	**0.4574**	**0.9718**	**-0.7271**	**0.775**	**0.8996**
CFAH_NGFYPATR	**339**	**0.8761**	**0.9756**	**0.4679**	**0.325**	**0.8166**
CFAI_IVIEYVDR	314	0.7761	0.9718	-0.518	0.709	0.8648
CLUS_ASSIIDELFQDR	394	0.2439	0.9718	1.3485	0.078	0.7176
CLUS_ELDESLQVAER	339	0.8761	0.9756	0.2241	0.403	0.8166
CLUS_FMETVAEK	392	0.2591	0.9718	1.105	0.137	0.78
CLUS_IDSLLENDR	386	0.3085	0.9718	1.182	0.097	0.7275
CO3_AGDFLEANYMNLQR	330	1	1	-0.3689	0.68	0.8446
CO3_GYTQQLAFR	284	0.4038	0.9718	-0.783	0.801	0.913

CO3_TGLQEVEVK	347	0.7621	0.9718	0.4683	0.319	0.8166
CO4A_DFALLSLQVPLK	**260**	**0.2022**	**0.9718**	**-1.5154**	**0.935**	**0.9546**
CO4A_VGDTLNLNLR	253	0.1604	0.9718	-1.5487	0.935	0.9546
CO4A_VLSLAQEQVGGSPEK	239	0.0968	0.9718	-2.0765	0.979	0.984
CO8A_AMAVEDIISR	369	0.4799	0.9718	0.6845	0.228	0.8166
CO8A_HTSLGPLEAK	352	0.6932	0.9718	0.4328	0.322	0.8166
CO8B_SGFSFGFK	283	0.3935	0.9718	-0.9423	0.825	0.9193
CO9_LSPIYNLVPVK	**325**	**0.9342**	**0.9947**	**-0.0363**	**0.51**	**0.8209**
CO9_VVEESELAR	328	0.978	0.9978	0.0487	0.455	0.8166
CXCL7_NIQSLEVIGK	334	0.9488	0.9947	0.4132	0.328	0.8166
DEMA_VTSNLGK	469	0.011	0.9718	2.5913	0.006	0.39
ENOA_TIAPALVSK	347	0.7621	0.9718	0.3635	0.35	0.8166
F13A_STVLTIPEIIIK	237	0.0896	0.9718	-1.5858	0.948	0.9628
FETUB_LVVLPFPK	286	0.4248	0.9718	-0.4454	0.679	0.8446
FIBA_GSESGIFTNTK	256	0.1774	0.9718	-1.2212	0.877	0.9276
FIBB_AHYGGFTVQNEANK	242	0.1084	0.9718	-1.3051	0.88	0.9276
FIBG_EGFGHLSPTGTTEFWLGNEK	227	0.06	0.9718	-1.8639	0.964	0.974
FINC_SYTITGLQPGTDYK	369	0.4799	0.9718	0.3373	0.39	0.8166
FINC_YSFCTDHTVLVQTR	383	0.3354	0.9718	0.9137	0.192	0.8166
GDIB_DLGTESQIFISR	334	0.9488	0.9947	0.0653	0.457	0.8166
GDIB_FVSISDLLVPK	305	0.653	0.9718	-0.1792	0.577	0.8446
GELS_AGALNSNDAFVLK	311	0.7343	0.9718	-0.3865	0.653	0.8446
GELS_SEDCFILDHGK	383	0.3354	0.9718	1.1538	0.109	0.767
H4_DAVTYTEHAK	321	0.8761	0.9756	-0.1626	0.562	0.843
H4_VFLENVIR	257	0.1834	0.9718	-1.3698	0.928	0.9546
HABP2_VVLGDQDLK	274	0.3085	0.9718	-1.0142	0.842	0.9224
HBG1_MVTAVASALSSR	439	0.0465	0.9718	2.0508	0.022	0.7176
HEMO_NFPSPVDAAFR	301	0.601	0.9718	-0.3326	0.628	0.8446
HEMO_VDGALCMEK	422	0.0932	0.9718	1.7543	0.033	0.7176
HEP2_FAFNLYR	340	0.8616	0.9756	0.2125	0.406	0.8166
HEP2_IAIDLFK	319	0.8472	0.9718	-0.4163	0.678	0.8446
HEP2_TLEAQLTPR	304	0.6398	0.9718	-0.3395	0.648	0.8446
HINT1_IIFEDDR	356	0.6398	0.9718	0.8128	0.208	0.8166
HPT_DYAEVGR	334	0.9488	0.9947	-0.022	0.515	0.8209
HPT_VGYVSGWGR	314	0.7761	0.9718	-0.4063	0.671	0.8446
HPT_VTSIQDWVQK	319	0.8472	0.9718	-0.5095	0.714	0.8648

HRG_GGEGTGYFVDFSVR	395	0.2366	0.9718	0.1863	0.421	0.8166
IC1_FQPTLLTLPR	342	0.8329	0.9718	0.2968	0.383	0.8166
IC1_LLDSLPSDTR	377	0.3935	0.9718	0.8756	0.19	0.8166
IC1_TNLESILSYPK	331	0.9927	0.9978	0.3165	0.352	0.8166
IF4B_SILPTAPR	**308**	**0.6932**	**0.9718**	**-0.414**	**0.662**	**0.8446**
IGHA1_TPLTATLSK	389	0.2831	0.9718	0.2727	0.401	0.8166
IGHG1_FNWYVDGVEVHNAK	**319**	**0.8472**	**0.9718**	**-0.1175**	**0.548**	**0.8414**
IGHG1_GPSVFPLAPSSK	323	0.9051	0.986	-0.1826	0.601	0.8446
IGHG3_DTLMISR	**351**	**0.7068**	**0.9718**	**0.328**	**0.348**	**0.8166**
IGHG3_NQVSLTCLVK	373	0.4355	0.9718	0.724	0.196	0.8166
IGHM_GFPSVLR	392	0.2591	0.9718	1.2503	0.092	0.7176
IGHM_QIQVSWLR	386	0.3085	0.9718	1.4108	0.068	0.7176
IGHM_YAATSQVLLPSK	396	0.2294	0.9718	1.4802	0.064	0.7176
ITIH1_AAISGENAGLVR	367	0.503	0.9718	0.8562	0.188	0.8166
ITIH4_GPDVLTATVSGK	367	0.503	0.9718	0.6036	0.261	0.8166
ITIH4_ILDDLSPR	355	0.653	0.9718	0.2693	0.384	0.8166
KAD1_IIFVVGGPGSGK	**354**	**0.6663**	**0.9718**	**0.3986**	**0.329**	**0.8166**
KNG1_DIPTNSPELEETLTHTITK	362	0.5633	0.9718	0.6483	0.237	0.8166
KNG1_TVGSDTFYSFK	314	0.7761	0.9718	-0.3518	0.662	0.8446
KNG1_YFIDFVAR	318	0.8329	0.9718	-0.2699	0.616	0.8446
KPYR_GDLGIEIPAEK	298	0.5633	0.9718	0.1268	0.436	0.8166
LEG3_IALDFQR	297	0.5509	0.9718	-0.1038	0.533	0.8315
LUM_SLEDLQLTHNK	317	0.8186	0.9718	-0.0161	0.504	0.8209
MUCB_GQPLSPEK	427	0.0766	0.9718	2.0602	0.026	0.7176
NDKA_DRPFFAGLVK	315	0.7902	0.9718	0.0855	0.444	0.8166
PARK7_ALVILAK	335	0.9342	0.9947	0.0124	0.478	0.8176
PARK7_DGLILTSR	**351**	**0.7068**	**0.9718**	**0.1982**	**0.422**	**0.8166**
PEBP1_LYEQLSGK	353	0.6797	0.9718	0.1846	0.414	0.8166
PEDF_ELLDTVTAPQK	289	0.4574	0.9718	-0.8731	0.788	0.9092
PEDF_TVQAVLTVPK	307	0.6797	0.9718	-0.7745	0.817	0.9156
PERM_IANVFTNAFR	332	0.978	0.9978	-0.3615	0.671	0.8446
PLMN_FVTWIEGVMR	345	0.7902	0.9718	0.1902	0.415	0.8166
PMGE_HYGALIGLNR	349	0.7343	0.9718	0.295	0.374	0.8166
PNPH_FEVGDIMLIR	309	0.7068	0.9718	-0.6123	0.729	0.8721
PNPH_VFGFSLITNK	358	0.6138	0.9718	0.0327	0.478	0.8176
PON1_LLIGTVFHK	363	0.5509	0.9718	0.048	0.518	0.8209

PPAC_IELLGSYDPQK	295	0.5267	0.9718	-0.6997	0.765	0.8933
PPIA_FEDENFILK	322	0.8905	0.9756	-0.0477	0.506	0.8209
PRDX1_ADEGISFR	319	0.8472	0.9718	-0.3415	0.637	0.8446
PRDX2_SVDEALR	362	0.5633	0.9718	0.8232	0.187	0.8166
PRDX2_TDEGIAYR	363	0.5509	0.9718	0.8321	0.19	0.8166
PRDX6_LSILYPATTGR	314	0.7761	0.9718	0.1549	0.439	0.8166
PROF1_TLVLLMGK	248	0.1348	0.9718	-0.9716	0.808	0.913
PRS6A_DAFALAK	365	0.5267	0.9718	1.0828	0.124	0.78
PRS8_FIGEGAR	395	0.2366	0.9718	1.5378	0.046	0.7176
PSA2_AANGVVLATEK	368	0.4914	0.9718	0.6721	0.238	0.8166
PSA6_HITIFSPEGR	**310**	**0.7205**	**0.9718**	**-0.1396**	**0.554**	**0.842**
PSB7_GTTAVLTEK	377	0.3935	0.9718	1.0463	0.144	0.78
RANG_FLNAENAQK	420	0.1005	0.9718	1.4115	0.076	0.7176
SAA4_EALQGVGDMGR	311	0.7343	0.9718	-0.338	0.614	0.8446
SAMP_IVLGQEQDSYGGK	308	0.6932	0.9718	-0.2809	0.627	0.8446
SH3L3_VYSTSVTGSR	322	0.8905	0.9756	-0.2711	0.584	0.8446
TBA4A_EIIDPVLDR	341	0.8472	0.9718	0.1236	0.443	0.8166
THBG_NALALFVLPK	293	0.503	0.9718	-0.715	0.753	0.8881
THIO_VGEFSGANK	417	0.1125	0.9718	1.7057	0.036	0.7176
THRB_ELLESYIDGR	272	0.2914	0.9718	-1.028	0.849	0.923
TPIS_FFVGGNWK	333	0.9634	0.9978	-0.1934	0.59	0.8446
TRFE_EGYYGYTGAFR	361	0.5757	0.9718	0.6503	0.252	0.8166
TRFE_MYLGYEYVTAIR	357	0.6268	0.9718	0.0381	0.467	0.8166
TSP1_GTLLALER	304	0.6398	0.9718	-0.3326	0.617	0.8446
UB2L3_IYHPNIDEK	332	0.978	0.9978	0.0505	0.462	0.8166
VTDB_HLSLLTTLSNR	369	0.4799	0.9718	0.6605	0.237	0.8166
VTDB_VLEPTLK	348	0.7481	0.9718	0.6175	0.246	0.8166
VTNC_DVWGIEGPIDAAFTR	361	0.5757	0.9718	0.0552	0.501	0.8209
VTNC_DWHGVPGQVDAAMAGR	352	0.6932	0.9718	-0.242	0.622	0.8446
WDR1_VFASLPQVER	309	0.7068	0.9718	-0.0818	0.522	0.8209
BMI	**283**	**0.3935**	**0.9718**	**-0.3307**	**0.647**	**0.8446**

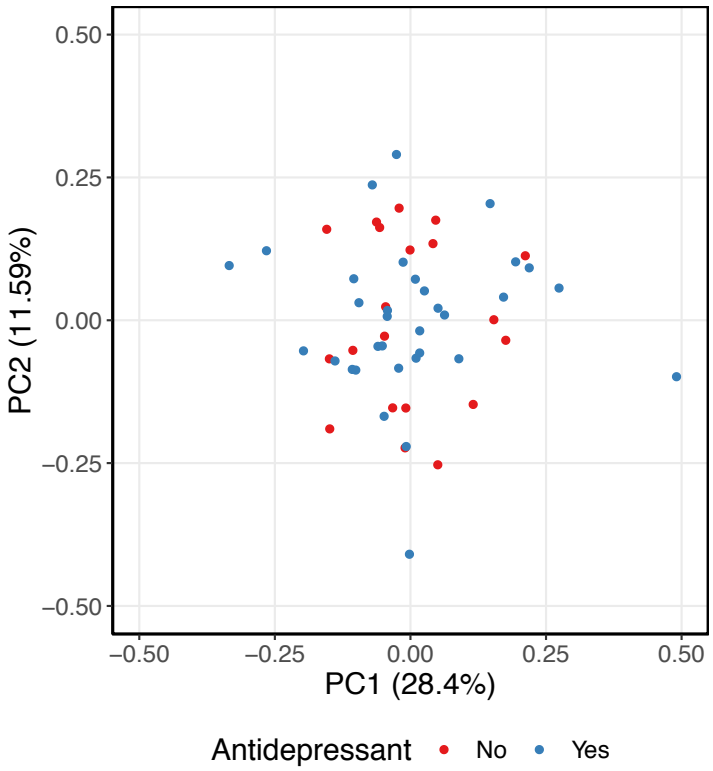


Figure 5.7. PCA plot of protein peptide abundance ratios of antidepressant users and non-users in the training set patient group (established current MDD patients).

The scores for the first two PCs are plotted with the percentage of variation accounted for by each PC shown in the axis labels. Data points are coloured by the patients' current antidepressant use (33 antidepressant users and 20 non-users). Abbreviations: MDD (major depressive disorder); PC (principal component); PCA (principal component analysis).

Table 5.4. Mann-Whitney *U* test results to assess the associations of the selected proteomic biomarkers with lifestyle factors.

The protein peptide abundances of the 16 biomarkers appearing the 15 prediction models were associated with binary measures of smoking, alcohol consumption, recreational drug use and poor self-rated physical health in 295 participants. The null hypothesis was that there was no statistically significant effect of the lifestyle factor on the abundance of the feature of interest. The *W*-value (Mann-Whitney *U* statistic), *p*-value and *q*-value (FDR-adjusted *p*-value) are shown for each test. *P*-values and *q*-values < 0.05 are shown in bold. Proteomic features are represented in a protein_peptide format, and their abbreviations can be found in the **Appendix** (**Table A. 1**). Abbreviations: FDR (false discovery rate); MDD (major depressive disorder).

Feature	Smoking			Alcohol			Recreational drug use			Poor physical health		
	W-value	*p*-value	*q*-value	*W*-value	*p*-value	*q*-value	*W*-value	*p*-value	*q*-value	*W*-value	*p*-value	*q*-value
A1AG1_SDVVYTDWK	22494	0.4524	0.857	26880	0.5322	0.7344	18105	0.1789	0.503	16999	**0.0006**	**0.0047**
A2GL_VAAGAFQGLR	22357	0.3962	0.857	27797	0.2136	0.6157	18798	0.0535	0.3054	18411	**0.0171**	0.0613
AL1A1_ILDLIESGK	25853.5	0.1033	0.6462	26716	0.6073	0.7752	15920	0.6139	0.8848	19654.5	0.1428	0.2978
APOA4_ISASAEELR	22260	0.359	0.857	28033	0.1606	0.6026	13667	**0.0159**	0.1612	21852	0.8737	0.9262
APOA4_LAPLAEDVR	21761	0.2038	0.7666	27754.5	0.2243	0.6216	12875	**0.0021**	0.1612	22033.5	0.7693	0.8519
APOE_LGPLVEQGR	25893	0.0975	0.6425	26138	0.9017	0.9576	17470	0.4198	0.7599	21179	0.7354	0.8307
CFAH_IDVHLVPDR	24339	0.5784	0.8932	28470	0.0895	0.5381	18414	0.1083	0.3965	14947	**0**	**0.0002**
CFAH_NGFYPATR	23338.5	0.8786	0.974	26955.5	0.4993	0.7229	17634	0.3444	0.6749	16732	**0.0003**	**0.0025**
CO4A_DFALLSLQVPLK	22970	0.6789	0.9275	27102	0.4386	0.6807	18449	0.102	0.3879	20662	0.4713	0.6773
CO9_LSPIYNLVPVK	24302	0.5965	0.897	28102	0.1471	0.6026	18807	0.0526	0.3054	20038	0.2373	0.4343
IF4B_SILPTAPR	22919	0.6526	0.9275	25774	0.9033	0.9576	16938	0.7216	0.9586	19455	0.1067	0.236
IGHG1_FNWYVDGVEVHNAK	26712.5	**0.0252**	0.6195	27439	0.3163	0.668	17862.5	0.2547	0.5814	23612	0.1441	0.2978
IGHG3_DTLMISR	26709.5	**0.0254**	0.6195	27588.5	0.27	0.668	17815	0.2719	0.6063	23598	0.147	0.3001

KAD1_IIFVVGGPGSGK	23664	0.9384	0.9948	25252	0.6355	0.7954	15337	0.3183	0.6639	22479	0.5335	0.7305
PARK7_DGLILTSR	24743	0.3997	0.857	25387	0.7018	0.8404	16252	0.8232	0.9906	20171	0.2786	0.4777
PSA6_HITIFSPEGR	25319	0.2112	0.7666	24872	0.465	0.6886	17552	0.381	0.7318	20537	0.4163	0.6261

5.4. Discussion

Despite the complexity and heterogeneity of MDD, many previous studies have used healthy controls as a reference group for comparison against patients, and investigated the group differences based on limited study participant information. This chapter not only examined individuals presenting with subclinical low mood as a more clinically relevant reference population, but also leveraged extensive participant information gathered through the Delta Study, including proteomic, clinical, sociodemographic/lifestyle and personality features. The convenience of DBS sampling, digital mental health assessment and telephone diagnostic interview enabled participants to complete all parts of the data collection remotely (from home or elsewhere, without the need to visit the clinic or laboratory). To the best of our knowledge, the Delta Study is the largest study to date involving a digital mental health assessment of mood disorders [381], [382].

Diagnostic prediction models were developed with the aim of identifying depressed patients among individuals presenting with low mood. This study demonstrated that the combined digital-biomarker prediction models were able to differentiate between new current MDD patients and low mood controls with a good and robust predictive performance (extrapolation test AUC = 0.80 ± 0.01), as well as between established non-current MDD patients and low mood controls with a fair to good, robust predictive performance (application AUC = 0.79 ± 0.01), despite having been trained to differentiate between established current MDD patients and low mood controls (training AUC = 0.94 ± 0.01). The decrease in model performance in comparison to the training performance could be seen as a result of the models having to go beyond their scope to make predictions on the extrapolation test set and the application set (given the differences in the patient groups), and rather than due to model overfitting, which was alleviated by the implementation of various statistical methods (LASSO regression, repeated nested cross-validation, feature extraction and model averaging). This generalisability suggests that a considerable part of the disease profile was commonly exhibited across the different patient groups of established current MDD patients, new current MDD patients and established non-current MDD patients. Although these findings are preliminary and require further validation on an independent dataset, they may represent an encouraging step towards the development of a diagnostic test for MDD.

A selection of proteomic biomarkers, sociodemographic and lifestyle factors and personality traits were found as important predictors of MDD. Several proteins were identified as biomarker candidates for MDD: MDD was associated with increased DBS levels of A1AG1, A2GL and CFAH, which have functional roles in immune/inflammatory response (anti-inflammatory response, acute-phase response, complement system); increased DBS levels of APOE, which is involved in lipid transport and metabolism; and decreased DBS levels of AL1A1, which is involved in retinol metabolism [359]. These findings are in agreement with previous reports of increased blood plasma levels of acute-phase proteins, including A1AG1 [383]–[385], and complement proteins, including CFAH [386]–[388], in MDD patients compared to controls, although the association of A2GL with depression has not been reported before. Thus, they are generally consistent with the observations that MDD patients exhibit activated inflammatory pathways, and support the role of immune system dysregulation in the pathophysiology of MDD [141], [150], [151]. It should be noted that A1AG1 and CFAH were also found to be significantly associated with poor self-rated physical health. Additionally, the APOE ε4 allele has been identified as a risk factor for MDD, especially among the elderly [389]–[391], and it has been suggested that the presence of the APOE ε4 allele may increase the vulnerability to depression in late life following stressful life events, including adverse childhood experiences (gene-environment interaction) [363], [364]. **Chapter 4** identified serum APOE as a biomarker for predicting future depression diagnosis in subthreshold symptomatic individuals. While AL1A1 oxidises retinaldehyde to retinoic acid (RA), the link between RA signalling and depression is supported by the evidence that brain areas implicated in RA signalling, such as the hippocampus, prefrontal cortex and striatum, overlap with those affected in depression [392].

Moreover, poor self-rated mental health was found as a strong and robust predictor of MDD, consistent with a previous report that those who rated their mental health as poor were more likely to have a first or recurrent depressive episode within the next year compared to those who rated their mental health as fair, good or excellent [393]. Others have demonstrated associations between poor self-rated mental health and increased psychological distress, increased utilisation of mental health and medical care services, and increased adherence to antidepressant treatment plans [394]–[397]. An association between MDD and increased BMI was also observed, which is a well-established phenomenon [366], [367] and consistent with the finding in **Chapter 4**. This is also in agreement with the findings that depression and obesity have an overlapping pathophysiology, including dysregulation of the HPA axis and

immune/inflammatory response [130]–[132]. In support of this, blood (serum or plasma) levels of the identified proteomic biomarkers of depression that are involved in immune/inflammatory response (*i.e.*, A1AG1, A2GL and CFAH) have been reported to be increased in obesity [398]–[401]. Furthermore, among various personality traits assessed using the NEO FFI, reduced daily experiences of positive emotions, such as happiness and joy, and greater empathy towards others (tender-mindedness) were found as important predictors of MDD. The association between MDD and personality has potentially important implications for elucidating the aetiology of MDD, identifying more homogeneous subgroups of disease, identifying at-risk individuals, tailoring treatment strategies, and predicting treatment response [402]. It should be noted that the combined digital-biomarker prediction models resulted in comparable yet improved predictive performance in relation to the digital-only prediction models, which demonstrates that the contribution of the digital features in predicting MDD was substantial, while that of the proteomic biomarkers was relatively modest.

An important aspect of the present study is the use of repeated nested cross-validation, which is essential for reliable model assessment, especially when large and representative datasets are not available [260]. Due to the study design, an adapted approach of repeated nested cross-validation was implemented, such that patients used for model assessment in the outer loop were never used for model selection in the inner loop. Fifteen prediction models were obtained from five times-repeated three-fold nested cross-validation, and despite using a robust model development approach (feature extraction and model averaging), some variation in feature selection and predictive performance was observed across the models. This indicates that the results were largely dependent on the random split of the dataset into training and test sets, which highlights the need for repeated nested cross-validation to allow for more informed decision-making in selecting and assessing prediction models [260]. Without using repeated nested cross-validation, any one of the 15 models could have been obtained, ignoring the possibility of the other models. As the random split was only relevant for the reference group (the patient groups remained constant), the observed variation in model selection was possibly due to heterogeneity in the low mood controls. Repeated nested cross-validation addressed this variation and improved model reproducibility by allowing the results across the different models to be pooled together to identify the most robust features and obtain an average estimate of model predictive performance.

Furthermore, this study clearly demonstrated the utility of DBS sampling for proteomic biomarker identification and its potential for application within a clinical setting. However, in spite of the many advantages offered by this sampling method, there were also challenges involved the quantitative analysis of DBS proteins compared to that of serum proteins. A potential major limitation was the quality of DBS samples, since the protein content could be affected by factors such as small spot size (*i.e.*, insufficient sample) and multiple spotting. Hence, the spotting of blood onto the DBS collection card was an additional step involved in DBS sample preparation in comparison to serum sample preparation which could potentially increase the variability [315]. As pilot studies from the group (unpublished) revealed that DBS sample quality could be improved by providing clear instructions for sample collection and submission, it was subsequently ensured that the DBS sample collection kits used in the present study contained clear instructions with regards to sampling, drying and posting of the DBS samples [296]. Yet, the possibility that the quality of the samples could be compromised in some cases due to participants' failing to carefully follow the instructions still remained [306]. Another additional step implemented to the DBS sample preparation workflow which could potentially increase the variability was the extraction of analytes from the sample collection card [315]. However, despite these concerns, it was demonstrated in a previous study from the group, *Ozcan et al. (2017)* [202], that the reproducibility of targeted protein quantification achieved in DBS was comparable to that achieved in serum. Results showed that the average variation in sample preparation (median CV) was 6.5% and 8.9% in serum and DBS samples, respectively, while the average biological variation (median CV) was 8.8% and 13.2% in serum and DBS samples, respectively. Additionally, a strong correlation was found between peptide abundances measured in DBS and serum samples which were collected from healthy volunteers, indicating relatively consistent measurements between the two sample formats.

There are several limitations to the present study, one of which is that prediction models were trained on established current MDD patients due to the limited availability of new current MDD patients in the dataset. The sample size of the training set was also small, particularly considering the large number of features examined in model selection. In addition, while the gender difference in the prevalence of MDD was reflected in the study participants, they could be considered as relatively young (average age of 26 years), given that the median age of onset of MDD is typically in the early to mid 20s [29]. Moreover, although it may be more practical to define a single risk prediction model of MDD for potential translation into clinical application, this was not feasible using the given dataset due to the demonstrated variation in feature selection and predictive performance across different models. It is expected that

performing the analysis on a larger, more representative and homogeneous dataset would not only enhance model reproducibility but also potentially allow for a single prediction model to be specified. Furthermore, the fact that a proportion of established current MDD patients were taking antidepressant medication at the time of data collection could potentially confound the results of the analysis, as the reported associations between MDD and the selected proteomic features could at least partially represent medication effects and not solely disease effects. However, this study demonstrated not only that there were no significant associations between the selected proteomic features and antidepressant medication use, but also that the prediction models comprised of these features showed good predictive performance on new current MDD patients who were not on antidepressant medication. Additionally, the potential confounding effects of the sociodemographic and lifestyle factors were accounted for by including them as predictors for model selection using group LASSO regression. Finally, despite the effort to conduct a comprehensive analysis of the various features that could be related to MDD, the possibility of disregarding other potentially relevant features should be acknowledged.

In conclusion, this chapter showed that currently depressed patients with a new MDD diagnosis could be identified among individuals presenting with low mood using risk prediction models comprised of DBS proteomic biomarkers, sociodemographic/lifestyle factors and personality traits. The same models could also be used to identify currently not depressed patients with an existing MDD diagnosis. An adapted approach of repeated nested cross-validation enabled an evaluation of the variation in feature selection and predictive performance that arised from choosing different splits of the data. This study also demonstrated the advantages of implementing DBS sampling for proteomic biomarker identification. These findings, although preliminary, may represent a promising step towards developing an objective, non-invasive and cost-effective diagnostic test for MDD in the future, which could potentially complement clinical evaluations, improve patient care, and reduce the overall socioeconomic disease burden. In order to develop a robust and accurate diagnostic test to be translated into routine clinical practice, particularly in primary care, extensive validation studies using large, representative and independent sample sets will need to be carried out.

Chapter 6 Final discussion and conclusions

The aim of this chapter is to provide an integrated summary of the major findings resulting from the work presented in this thesis and their significance. The limitations of the studies and the ways in which they could be addressed in future work are subsequently discussed.

6.1. Summary

The study presented in **Chapter 3** explored the utility of a multimodel-based approach to predict the probability of having a complex psychiatric disorder using high-dimensional biomarker data. A novel approach combining feature extraction and model averaging was developed to address model overfitting and model selection uncertainty, which are statistical problems that can limit the ability of prediction models to make accurate and reproducible predictions. A serum MS dataset (147 peptides representing 77 proteins) comprising of first-onset drug-naïve schizophrenia patients and healthy controls was used to demonstrate proof-of-concept. LASSO regression with ten-fold cross-validation was repeatedly applied on the training set to produce a set of 100 prediction models. In the presence of model selection uncertainty, feature extraction and model averaging were applied across the entire set of models to define a parsimonious weighted average prediction model that consisted of the most robust (*i.e.*, frequently selected) features and their weighted average coefficient estimates. The resulting model demonstrated the advantage of the multimodel approach over the conventional approach based on a single 'best' model in producing reproducible predictions of the probability of having a complex psychiatric disorder (AUC = 0.77), despite the potential for improvement. This multimodel approach was subsequently applied in **Chapters 4 and 5** to develop diagnostic prediction models of MDD.

The study presented in **Chapter 4** investigated the prediction of future onset of depression in subthreshold symptomatic individuals using their serum proteomic, sociodemographic and

clinical data (146 peptides representing 77 serum proteins, 22 sociodemographic factors and 30 clinical features). Diagnostic prediction models were trained to differentiate between first-episode MDD patients and subthreshold individuals who did not develop MDD within four years. The multimodel approach combining feature extraction and model averaging was applied to obtain a parsimonious diagnostic prediction model consisting of six serum proteins (AACT, APOE, APOH, FETUA, HBA and PHLD), three sociodemographic factors (BMI, childhood trauma and education level), and three depressive symptoms (sadness, fatigue and leaden paralysis). When the model was subsequently applied to differentiate between subthreshold individuals who developed MDD within two or four years and subthreshold individuals who did not develop MDD within four years, which involved going beyond the scope of the model, it achieved a fair predictive performance (AUC = 0.75).

The study presented in **Chapter 5** investigated the prediction of MDD in individuals presenting with subthreshold depression (subclinical low mood) using data from DBS proteomics (194 peptides representing 115 DBS proteins) and a novel digital mental health assessment (102 sociodemographic, clinical and personality characteristics). Repeated nested cross-validation was used to evaluate variation in model selection and predictive performance arising from choosing different splits of the data and ensure model reproducibility, and feature extraction and model averaging were applied in each round of nested cross-validation. Diagnostic prediction models that were trained to differentiate between currently depressed individuals with an existing MDD diagnosis (established current MDD patients) and low mood controls demonstrated a good predictive performance when extrapolated to differentiate between currently depressed individuals with a new MDD diagnosis (new current MDD patients) and low mood controls (AUC = 0.80 ± 0.01), as well as between currently not depressed individuals with an existing MDD diagnosis (established non-current MDD patients) and low mood controls (AUC = 0.79 ± 0.01). DBS proteins A1AG1, A2GL, AL1A1, APOE and CFAH were identified as predictors of MDD, which indicate immune system dysregulation. Poor self-rated mental health, BMI, reduced daily experiences of positive emotions and tender-mindedness were also identifed as predictors of MDD.

6.2. Significance of major findings

6.2.1. Study design and statistical methods for reproducible prediction model development

The studies presented in **Chapters 4 and 5** represent rigorous efforts, both in terms of design and methodology, to develop and validate clinical risk prediction models based on protein biomarkers and other non-biological patient characteristics, which can potentially act as a diagnostic aid for MDD. This was an important consideration as poor study design and incorrectly used statistical techniques have been frequently recognised as factors contributing to the lack of reproducible biomarker findings and the limited translation of biomarkers into clinical tests, despite numerous reports of significant candidate biomarkers in discovery studies [15], [261] (as discussed in **Chapter 1**). The studies in **Chapters 4 and 5** attempted to address and overcome some of the limitations in previous biomarker research by not only defining stringent criteria for sample inclusion and exclusion based on clear clinical objectives, but also applying a range of statistical methods, including the multimodel approach developed in **Chapter 3**, to identify features that are robustly associated with MDD and ensure that the resulting models are able to make accurate and reproducible predictions of the probability of having or developing MDD.

Whereas many diagnostic biomarker studies typically use healthy controls as a reference group for comparison against patients, the studies in **Chapters 4 and 5** examined individuals presenting with subthreshold depression to investigate a more clinically relevant and appropriate population for developing and validating diagnostic prediction models of MDD. The rationale behind this was that distinguishing MDD patients from individuals with subthreshold depression would better represent a situation that GPs would encounter relative to distinguishing MDD patients from healthy individuals, since it is more likely for those who seek help in primary care to be presenting with at least subthreshold levels of depressive symptoms than with no symptoms [19], [58]. An additional challenge to model selection and classification was provided by the reference groups being more similar to the patient groups. Therefore, a notable advantage of these studies is that the findings may be more generalisable to primary care settings than those obtained from studies that have been conducted using healthy controls. Moroever, **Chapters 4 and 5** investigated the group differences based on extensive information of individual participants gathered through the NESDA and the Delta

Study, respectively. The study in **Chapter 4** measured 198 features (146 peptides representing 77 proteins, 22 sociodemographic factors and 30 clinical characteristics) across 209 participants, while the study in **Chapter 5** measured 296 features (194 peptides representing 115 proteins, and 102 sociodemographic, clinical and personality characteristics) across 295 participants. The Delta Study (**Chapter 5**) was not only, to our knowledge, the largest study thus far which exploited a digital mental health assessment of mood disorders, but also designed in such a way as to enable all parts of the data collection (*i.e.*, DBS sampling, digital mental health assessment and telephone diagnostic interview) to be completed by the participants remotely without needing to visit the clinic or laboratory. In addition, the studies in **Chapters 4 and 5** investigated blood-based proteomic biomarkers in serum and DBS samples. While the use of serum sampling for blood-based biomarker discovery is widespread, the demonstration of the utility of DBS sampling, which is a relatively novel sampling technique in the field that has a clear potential for application in clinical practice, was particularly notable [310].

Furthermore, results of model selection in **Chapters 3, 4 and 5** consistently demonstrate the presence of model selection uncertainty (*i.e.*, the sensitivity of model selection to small changes in the data), which is often overlooked in clinical prediction studies despite its significant impact on model reproducibility and generalisability (discussed in **Chapters 1 and 3**). This highlights the importance of implementing appropriate statistical techniques that allow for model selection uncertainty to be properly assessed and accounted for, especially considering that it may be realistically difficult to overcome the pre-analytical factors which contribute to it, such as small sample size and sample heterogeneity. In relation to this, the work presented in **Chapter 3** illustrates an effort to explore and develop a novel statistical approach that combines model averaging with feature extraction to obtain reproducible predictions of the probability of having a complex psychiatric disorder. While LASSO regression allows for sparse multivariable prediction models to be obtained by simultaneously performing regularisation and variable selection [281], the uncertainty involved the process of model selection was evaluated by repeatedly applying LASSO regression with ten-fold cross-validation on the training set to obtain a set of 100 models. Relative model weights (Akaike weights) were subsequently calculated for each model based on the AIC_c and used to assess the relative strength of the model among the candidate set of models. The relative importance of the features was assessed by measuring their selection fractions (the proportion of models out of 100 in which a particular feature was selected) and selection probabilities (the probability

that a particular feature was a component of the best model, calculated using Akaike model weights). In the presence of model selection uncertainty, defined as there being less than 90% chance that the model with the highest weight was the best model approximating the outcome given the candidate set of models, a multimodel approach was adopted to produce inferences and predictions that were derived across the full set of models rather than conditional on a specific model. This was important, particularly given the considerable variability in feature selection across the 100 models, as the risk of selecting one of the less probable models by chance could be reduced [294]. Feature extraction enabled the most important set of features to be identified for the average model and less important ones to be excluded, further alleviating model overfitting and enhancing model interpretability, while model averaging enabled the coefficient estimates of the identified features to be averaged across the full set of models in a weighted manner. Among the four weighted average prediction models analysed (defined by applying feature extraction using selection fraction and probability thresholds of 0.9 and 0.8, followed by model averaging), the most reproducible predictions of the probability of having a complex psychiatric disorder were obtained by the simplest weighted average model consisting of six features with selection fractions ≥ 0.9 (training AUC = 0.81; test AUC = 0.77). Increasing model complexity had the effect of increasing overfitting, as expected. Thus, the study in **Chapter 3** demonstrates that the combined use of feature extraction and model averaging could effectively address model selection uncertainty and model overfitting when analysing high-dimensional biomarker data, and thereby, result in a parsimonious prediction model with improved generalisability. It should be noted that, while model averaging using Akaike weights is an established method for deriving inferences and predictions from a set of models in the presence of model selection uncertainty [284], [289], [290], the novelty of the multimodel approach specifically lies in the generation of 100 models from the repeated application of the model selection method (*e.g.*, LASSO regression with ten-fold cross-validation) on the training set; the assessment of relative feature importance by measuring the proportion of models out of 100 in which a particular feature was selected (selection fraction); and feature extraction to identify the most important features (features with selection fractions ≥ 0.9), which was critical for obtaining a parsimonious and generalisable prediction model [294]. Obtaining a parsimonious prediction model, which achieved sufficient classification performance with as few features as possible, was important to not only enhance the interpretability of the findings, but also improve convenience and cost-effectiveness considering its potential use in a clinical setting. These considerations were taken forward to prediction model development in **Chapters 4 and 5**, such that the models were assessed by

their abilities to produce accurate and reproducible predictions of having or developing MDD using the smallest feasible set of features. The applications of the multimodel approach (feature extraction using selection fraction threshold 0.9, followed by model averaging across the set of models) in **Chapters 4 and 5** resulted in parsimonious and generalisable diagnostic prediction models of MDD. Therefore, the findings across **Chapters 3, 4 and 5** collectively demonstrate the utility of the multimodel approach in producing reproducible predictions of a complex psychiatric disorder, not only when the clinical conditions of the investigated participants were distinct (such as in **Chapter 3** where first-onset drug-naïve schizophrenia patients were compared against healthy controls), but also when the clinical conditions of the investigated participants were less distinct and more reflective of the relevant clinical population (such as in **Chapters 4 and 5** where MDD patients were compared against individuals with subthreshold depression).

In **Chapters 3, 4 and 5**, protein quantification was achieved by the selected measurement of surrogate peptides using MRM-MS [210]. The use of surrogate peptides for the quantification of target peptides is a well-established and widely used method; therefore, differential expression of a peptide was considered to represent differential expression of the corresponding protein (*i.e.*, protein of interest). Each target protein was represented by between one and five peptides, and the abundance levels of the individual peptides were used for statistical analysis to determine whether or not they were differentially expressed between the clinical groups of interest. The identification of individual peptides as candidate biomarkers (*i.e.*, important predictors of disease) indicates that it was the abundance levels of the peptides within their corresponding proteins, and not necessarily those of the whole proteins, that were altered. Similarly, there were other peptides which represented the same proteins but were not identified as candidate biomarkers. This could be because different peptides representing the same protein are likely to be highly correlated with each other, and the nature of the LASSO method is such that when there is a set of highly correlated features, only one feature from the set tends to be selected [335].

6.2.2. Clinical implications of findings

6.2.2.1. Prediction models

In both **Chapters 4 and 5**, the diagnostic prediction models had to make predictions beyond their scopes, as the patient groups comprising the test sets (extrapolation test and/or application sets) differed to those comprising the training sets. In **Chapter 4**, the training set patient group consisted of first-episode MDD patients (n = 86), whereas the extrapolation test set patient group consisted of subthreshold individuals who developed MDD within two or four years (n = 37); the shared reference group consisted of subthreshold individuals who did not develop MDD within four years (n = 86). In **Chapter 5**, the training set patient group consisted of established current MDD patients (n = 53), whereas the extrapolation test set patient group consisted of new current MDD patients (n = 40), and the application set patient group consisted of established non-current MDD patients (n = 72); the reference group consisted of low mood controls (n = 130). Therefore, although the model predictive performance was reduced relative to the training performance (**Chapter 4** training AUC = 0.94, extrapolation test AUC = 0.75; **Chapter 5** training AUC = 0.94, extrapolation test AUC = 0.80, application AUC = 0.79), this can be considered to be a result of the models having to go beyond their scopes to make predictions on the extrapolation test and/or application sets, rather than due to limited model generalisability, as a range of statistical methods (LASSO regression, repeated K-fold cross-validation, feature extraction and model averaging) were purposefully applied to reduce model overfitting and account for model selection uncertainty. This generalisability suggests that a considerable part of the disease profile, as represented by a combination of proteomic biomarkers and sociodemographic and other patient characteristics, was shared across the different patient groups of first-onset MDD patients and subthreshold individuals who developed MDD in **Chapter 4**, and established current MDD patients, new current MDD patients, and established non-current MDD patients in **Chapter 5**. In particular, the strength of the study in **Chapter 4** is that the disease profile of MDD was characterised based on first-onset drug-naïve MDD patients. Although in **Chapter 5**, the disease profile of MDD was characterised based on established current MDD patients, a proportion of whom were taking antidepressant medication, the findings that there were no significant associations between the selected proteomic features and the participants' antidepressant use, and that the prediction models showed a good predictive performance on new current MDD patients who were not on

antidepressant medication, demonstrated that the characterised disease profile was unlikely to be confounded by the effects of antidepressant medication.

Moreover, results of **Chapters 4 and 5** highlight the importance of combining different aspects of patient data for developing clinically useful disease prediction models. Blood (serum or DBS) proteomic biomarkers, as well as sociodemographic/lifestyle factors and other patient characteristics, were identified as important predictors of MDD outcome. These findings are consistent with the idea that, due to the complex and heterogeneous nature of MDD, the disease characteristics are unlikely to be sufficiently captured by any single biomarker or feature, and instead, a panel comprised of multiple biomarkers and features would be needed to achieve sufficiently high discriminatory power for clinical application [14], [223], [236], [237] (as discussed in **Chapter 1**). In **Chapter 4**, the combined biomarker-sociodemographic-clinical model achieved an improved predictive performance for predicting future MDD onset among individuals with subthreshold depression in relation to both the biomarker-only model and the sociodemographic-clinical-only model (extrapolation test AUC: combined = 0.75, biomarker-only = 0.70, sociodemographic-clinical-only = 0.67). Additionally, in **Chapter 5**, the combined digital-biomarker models achieved a comparable yet improved predictive performance for predicting MDD outcome among individuals with subthreshold depression in relation to the digital-only models, and a significantly improved predictive performance in relation to the biomarker-only models (extrapolation test AUC : combined = 0.80; biomarker-only = 0.56; digital-only = 0.79; application AUC: combined = 0.79; biomarker-only = 0.63; digital-only = 0.76). Thus, in **Chapter 4**, the contribution of the serum proteomic biomarkers in predicting MDD on the extrapolation test set was comparable to, yet greater than, that of the non-biological features (*i.e.*, sociodemographic factors and depressive symptoms), whereas in **Chapter 5**, the contribution of the digital features in predicting MDD on the extrapolation test and application sets was substantially greater than that of the DBS proteomic biomarkers.

It should be noted that the biomarker findings are not directly comparable between **Chapters 4 and 5** as different yet partially overlapping panels of protein peptides were employed for MRM-based targeted protein quantification in serum and DBS. Not all protein peptides could be measured in both serum and DBS samples, the main difference between the serum and DBS proteomes being that the latter contained additional proteins derived from red and white blood cells [403], [404]. This could account for the lack of overlap in the protein peptides that were identified as biomarker candidates between the two studies. While 146 peptides representing

77 serum proteins were investigated in **Chapter 4** and 194 peptides representing 115 DBS proteins were investigated in **Chapter 5**, 89 peptides representing 51 proteins were investigated in both **Chapters 4 and 5**; 57 peptides representing 45 proteins were investigated in **Chapter 4** only; and 105 peptides representing 86 proteins were investigated in **Chapter 5** only. Consequently, among the six serum protein peptides identified as predictors of MDD in **Chapter 4**, only three (AACT_ADLSGITGAR, APOE_ALMDETMK and APOH_EHSSLAFWK) were investigated in DBS in **Chapter 5**; and among the five DBS protein peptides identified as predictors of MDD in **Chapter 5**, only one (APOE_LGPLVEQGR) was investigated in serum in **Chapter 4**.

Despite the need for further validation, these findings demonstrate the potential of such prediction models to be used as an aid to the detection and diagnosis of MDD in clinical practice. In particular, the findings in **Chapter 4** suggest that it may be possible to detect disease indications in subthreshold individuals up to four years prior to diagnosis, which has important clinical implications regarding the identification of high-risk individuals early in the disease course. Targeting these individuals with indicated preventive interventions could be effective in delaying or even preventing the onset of MDD [78]–[80] and further contribute to the reduction of the disease burden.

6.2.2.2. Predictors of MDD and clinical interpretations

Overall, although further investigations are required, the findings in **Chapters 4 and 5** collectively point to the presence of an altered immune/inflammatory profile in the pathophysiology of depression: many of the proteomic biomarker candidates identified in **Chapters 4 and 5** are involved in immune/inflammatory processes. MDD was found to be associated with elevated levels of acute-phase proteins, such as serum AACT and FETUA in **Chapter 4** and DBS A1AG1 and A2GL in **Chapter 5**, as well as with elevated levels of the complement factor CFAH in DBS in **Chapter 5**. These findings are in agreement with the extensive literature showing that depression is associated with activated inflammatory pathways, as manifested by increased levels of pro-inflammatory cytokines, acute-phase proteins, chemokines and cellular adhesion molecules, and support the role of immune system dysregulation in the pathophysiology of depression [141], [150], [151]. Thus, the findings in **Chapters 4 and 5** are consistent with the notion that depression is a systematic disorder and that the disease-related alterations can be traced in the peripheral system [14]. However, a common complication in biomarker studies is that it remains to be established whether the

observed alterations are involved in the aetiology of MDD or represent an adaptive response to the pathophysiological changes [14]. Given the evidence that exposure to stressful stimuli can induce inflammation both in the brain and peripherally [405], [406], the possibility that the observed upregulation of the immune/inflammatory response represents pathophysiological alterations that are associated with a stress response rather than specifically with MDD should be acknowledged. Nevertheless, the joint effects of proteomic features as well as other patient characteristics to differentiate between MDD patients and subthreshold individuals were investigated in **Chapters 4 and 5**, and the set of biomarkers identified consisted of not only of inflammatory proteins, but also other proteins that are not known to be involved in stress response. Thus, the combination of the different proteins (and the sociodemographic/lifestyle factors) could be considered to represent a disease-specific profile of MDD as opposed to a stress response.

A consistent finding across **Chapters 4 and 5** was that MDD patients had higher BMI compared to individuals with subthreshold depression. While the relationship between depression and obesity (defined as BMI (kg/m^2) ≥ 30) is an established phenomenon that has been repeatedly observed [366], [367], [407], a bidirectional association between depression and obesity was also reported in a meta-analysis of longitudinal studies, such that not only were obese individuals 55% more likely to develop depression, but depressed individuals were also 58% more likely to become obese [131]. In addition, the positive association between depression and BMI was found to be more notable among women, which may reflect the increased incidence and prevalence of depression in women [407]. Importantly, depression and obesity share alterations in immune/inflammatory and neuroendocrine pathways [130]–[132], [408], [409], and this is further supported by the evidence that blood levels of the inflammatory biomarkers of depression identified in **Chapter 5** (A1AG1, A2GL and CFAH) have been found to be increased in obesity [398]–[401]. Moreover, childhood trauma was identified as a strong predictor of MDD in **Chapter 4**. Adverse childhood experience is a well-known risk factor for developing depression in adulthood, particularly in response to stress [115]–[117], [410]. This relationship has been proposed to be mediated by dysregulation of the HPA axis and inflammatory pathways, as childhood trauma has been associated with several of the characteristic pathophysiological features of depression, including sensitisation of the neuroendocrine stress response, CRF hyperactivity, glucocorticoid resistance and immune activation [139], [146] (as discussed in **Chapter 1**). Additionally, a positive association between childhood trauma and obesity has been observed [411]–[413]. These relationships

could potentially be explained by the idea that childhood adversity leads to emotional dysregulation through activation of the stress response; in turn, emotional dysregulation has been associated with an increased appetite and a preference for and consumption of foods high in sugar and fat ('emotional eating'), which are observed in depression and could lead to excessive weight gain [130], [414]–[417]. Therefore, these findings support potential disturbances in the immune/inflammatory and neuroendocrine pathways in depression. While alterations in immune/inflammatory response is clearly supported by the biomarker findings in **Chapters 4 and 5**, this is not necessarily the case for the HPA axis, as none of the identified biomarker candidates are known to be directly involved in the neuroendocrine system. Nevertheless, the observed associations between MDD, BMI and childhood trauma could be taken as implicative of not only a dysregulated immune/inflammatory system but also a dysregulated HPA axis. It has been suggested that interactions may occur between these pathways and that there may be a common pathophysiological mechanism linking them, whereby activation of immune/inflammatory responses in MDD could lead to HPA axis hyperactivity, and vice versa [14], [141], [150], [160].

In addition, preliminary evidence suggests that the relationship of depression with HPA axis hyperactivity and immune activation may be mainly driven by somatic symptoms, such as fatigue and sleeping problems; on the other hand, the role of cognitive symptoms, such as sadness and anhedonia, may be considerably smaller [418]–[422]. Thus, elevated inflammatory and neuroendocrine responses may contribute to the development of the somatic components of depression [421], which is consistent with the findings of the somatic symptoms of fatigue and leaden paralysis as strong predictors of MDD outcome in **Chapter 4** (but not sadness which was found as another strong predictor). The differential association of somatic and cognitive symptoms of depression with HPA axis hyperactivity and immnue activation provides insight into the heterogeneity of depression and highlights the importance of considering the effects of specific symptom profiles on the pathophysiological mechanisms of depression and the therapeutic efficacy of antidepressants [422]. This also suggests that the current diagnostic strategy, which considers individuals symptoms as interchangeable and relies on summing them together to establish diagnoses, could result in a loss of information [371], [372], [423].

Furthermore, the findings appear to be consistent with the diathesis-stress model for depression, which proposes that the onset of depression is a result of an interaction between an individual's vulnerability (*i.e.*, diathesis) and environmental stress (caused by life experiences) [106], [424].

As genetic risk factors can be regarded as a genetic diathesis, the interaction between genetic diathesis and stress can be considered as a gene-environment interaction [425]. While stressful life events have been consistently recognised as a strong determinant of depressive symptoms and the onset of MDD [118], [426]–[428], according to the diathesis-stress model, individuals develop depression if the combined effect of latent diathesis and stress exceeds a certain threshold, whereby an individual with a greater vulnerability requires a smaller amount of stress to trigger the manifestation of symptoms [424], [429]. This is supported by the finding from a study of female-female twin pairs that the onset of MDEs could be predicted by the interaction between genetic risk factors and stressful life events [427], and explains why some individuals develop depression following stressful life events whereas others never do so. In addition, the diathesis-stress model proposes that the interaction between diathesis and stress increases the liability to depression beyond what would be expected from their additive contributions. Consistent with this, recent studies have shown that, due to such gene-environment effects, those with an inherent genetic predisposition to MDD and a high number of recent stressful life events are at an extra risk of depression, although larger sample sizes may be required to confirm these findings [425], [429]. Moreover, it has been suggested that as different individuals have various levels of inherent vulnerability and experience various levels of stress in their lives, diathesis-stress interactions may result in a fluctuating level of depressive symptoms across the population [430]. This could be consistent with the spectrum view of depression, which proposes that depressive disorders exist along a dimensional continuum of symptomatic severity, and account for the clinical observations of subthreshold depression, partial remission and residual symptoms [430].

MDD is a complex and multifactorial disorder that arises from complex interactions between genetic risk factors, molecular and functional disturbances and environmental exposures [48], [106]. From a developmental viewpoint, genetic predispositions and childhood adversity may result in sensitisation of the neuroendocrine stress response system involving the HPA axis, possibly through epigenetic modifications (*i.e.*, gene-environment interaction), which can, in turn, lead to immune system activation, increase the vulnerability to depression, as well as promote unhealthy lifestyles and obesity in later life [168], [172], [173], [431]. The identification of APOE, a protein involved in lipid transport and metabolism, as a biomarker candidate for MDD in both **Chapters 4 and 5** may represent a potential interaction between genetic and environmental factors. Previous studies have revealed an association between APOE polymorphism and depression, in particular late-life depression [389]–[391], and it has

been suggested that the APOE ε4 allele may modulate the adverse effects of stress, including childhood adversity, and lead to an increase the vulnerability to depression in later life [363]. Additionally, the APOE ε4 genotype has been associated with an increased severity of depression among depressed patients [432], as well as with stronger responses to stress [364]. However, it should be noted that APOE levels were reduced in MDD patients compared to subthreshold individuals in **Chapter 4**, whereas they were increased in MDD patients compared to subthreshold individuals (low mood controls) in **Chapter 5**. Such inconsistency could be partially accounted for by the use of different sample materials (serum in **Chapter 4** and DBS in **Chapter 5**), as well as by the identification of different peptides representing APOE (ALMDETMK in **Chapter 4** and LGPLVEQGR in **Chapter 5**; only the abundance levels of the peptides within the proteins, and not necessarily those of the whole proteins, were altered).

6.3. Limitations and future work

There are several limitations to the work presented in this thesis. The sample sizes of the datasets used in **Chapters 3, 4 and 5** were relatively small, particularly considering the large numbers of features analysed in model selection. Given the general difficulty associated with recruiting appropriate patient and reference samples, small sample size is a major limitation in many psychiatric studies. Nevertheless, the multimodel approach developed in **Chapter 3** and applied in **Chapters 4 and 5** was effective in accounting for model selection uncertainty and alleviating model overfitting, which can limit model reproducibility when model selection is performed on high-dimensional data (small n, large p). Additionally, in **Chapters 4 and 5**, prediction models had to be extrapolated to the target sample groups of subthreshold individuals who developed MDD and new current MDD patients, respectively, after having been trained on first-onset MDD patients and established current MDD patients, due to the limited availability of the target sample groups in the datasets. The accuracy and reproducibility of the prediction models could potentially be improved if they could be both trained and tested on the relevant target sample groups of sufficient sample sizes. Furthermore, although the models were validated on independent datasets (*i.e.*, datasets other than those used for model training), it would be essential to also validate them on independent cohorts of patient and reference samples recruited from different clinical centres in order to further determine the

generalisability of the models. This was not feasible as no such cohorts were available. Hence, extensive studies using large, representative and independent sample sets will need to be conducted to develop robust, accurate and reliable prediction models to aid in the clinical diagnosis of MDD, particularly within the primary care setting. Moreover, despite potential gender differences in proteomic biomarker profiles of MDD, these were not investigated within the current scope of the thesis. Further analyses could therefore be carried out on the existing datasets to investigate differences in serum and/or DBS biomarkers of MDD between males and females as well as gender-specific associations between MDD and biomarker levels, as reported in previous studies [433], [434]. These would help to not only elucidate gender differences in the pathophysiology of MDD but also explore their potential utility in the development of a diagnostic biomarker test. The existing datasets could also be used to examine the effects of various sociodemographic and lifestyle factors, such as smoking, alcohol abuse and childhood trauma, on proteomic biomarker profiles, which were not explicitly studied within the present thesis. As these sociodemographic and lifestyle factors may potentially interact with the biomarkers in predicting the disease outcome, it would be important to explore these interactions in future analyses and enable any significant interactions to be taken into account for disease classification and prediction.

As the findings of **Chapters 4 and 5** reveal that MDD patients could be differentiated from individuals presenting with subthreshold levels of symptoms, further studies need to be conducted in subthreshold individuals for a better characterisation of the condition. This would allow for earlier and more accurate diagnosis of MDD, and thereby result in earlier interventions and improved patient outcomes. In particular, given the growing evidence that subthreshold depression may represent a prodromal stage of MDD [55], [56], [358], it would be of interest to examine whether diagnostic (or prognostic) prediction models could identify individuals at different stages of the prodromal stage and/or investigate the role of proteomic biomarkers in the progression from subthreshold depression to full-blown MDD. In addition, given the heterogeneity of depression and the discussed differential association of somatic and cognitive symptoms of depression with HPA axis hyperactivity and immnue activation, it may be relevant for future studies to identify subgroups of MDD patients based on their symptom profiles (*i.e.*, those who present mainly with somatic symptoms and those who present mainly with cognitive symptoms) and develop separate diagnostic prediction models for each subgroup. This may potentially improve the accuracy and reproducibility of the outcome predictions. Future studies should also assess whether the diagnostic biomarkers and patient

characteristics that have been identified in this work are specific for MDD or whether they are also altered in other psychiatric disorders. In other words, the differential diagnostic potential of prediction models need to be evaluated, especially in differentiating between MDD and BD, since BD is frequently misdiagnosed as MDD and the two disorders share overlaping depressive symptoms [228].

Furthermore, the contribution of DBS proteomic biomarkers to the prediction of MDD was limited relative to that of the digital features in **Chapter 5**, even though serum proteomic biomarkers made a considerable contribution to the prediction of MDD in **Chapter 4**. While the overall precision of a protein quantification assay depends on the reproducibility of sample preparation as well as that of MRM-MS analysis, two additional steps were involved in DBS sample preparation which could potentially increase the variability when compared with serum sample preparation; these were the spotting of blood onto the DBS collection card and the extraction of analytes from the card [315]. In particular, failure of the participants to follow instructions with regards to sampling, drying and posting could lead to poor quality of DBS samples [306]. Therefore, better controlling for these steps in the sample collection and preparation procedures in future studies could enhance the precision and accuracy with which proteins are quantified in DBS samples, and potentially allow for more important and robust DBS-based biomarkers to be identified.

It should be noted that while childhood trauma was investigated as a potential predictor of MDD in both **Chapters 4 and 5**, it was only identified as a predictor in **Chapter 4**. This is expected to be not only due to differences in the datasets analysed, but also at least partially due to difference in the way in which childhood trauma was assessed and measured in the respective clinical studies. In **Chapter 4** (NESDA), childhood trauma was assessed using the childhood trauma index, where participants were asked about their experiences of emotional neglect, psychological abuse, physical abuse and sexual abuse in early life and sum scores ranging from zero to eight were calculated to take into account the number of types and frequency of childhood trauma experienced [354]. On the other hand, in **Chapter 5** (Delta Study), participants were only asked whether or not they had had any emotionally painful or distressing experiences in early childhood, resulting in a binary (yes/no) variable. This highlights the potential importance of measuring a predictor in a way that makes the most out of the information contained, which may be a valuable consideration in designing studies in the future. Moreover, as much of the patient information collected in the NESDA and the Delta

Study in **Chapters 4 and 5**, including sociodemographic, lifestyle, clinical and/or personality features, was assessed by self-report, these variables may be inaccurate and potentially suffer from social desirability response bias (that is, the tendency for participants to present a favourable image of themselves on questionnaires) [435].

Finally, despite the efforts to conduct a comprehensive analysis of the various features that could be associated with MDD in **Chapters 4 and 5**, it should be acknowledged that other potentially relevant features may have been overlooked. In particular, as the selection of proteins which were investigated by targeted MRM-MS analysis was based on their previous associations with psychiatric disorders as gathered from the scientific literature, it is possible that potentially important discriminatory proteins have not been included. Therefore, it would be essential to conduct discovery proteomic analysis using an untargeted approach such as shotgun proteomics to identify additional blood-based protein biomarker candidates for MDD [436]. This could be performed using a LC system coupled with a quadrupole time-of-flight (Q-TOF) mass spectrometer [437]. A potential challenge in discovery proteomic analysis is that, while biomarker candidates are generally likely to be present within the low concentration range, low abundance proteins tend to be masked by high abundance proteins, such as albumin, immunoglobulins and complement factors, which account for the majority of the total protein content [217] [300]. Therefore, removal of these high abundance proteins from the blood samples using methods such as immunodepletion in order to enhance the coverage of the blood proteome would be an essential consideration for the identification novel biomarker candidates [299], [301], [302]. Note that this was not necessary in targeted MS analysis in **Chapters 3, 4 and 5** as the proteins of interest were already known. Additionally, proteins encoded by genes which have been implicated in MDD based on findings from genomic studies should be investigated for their diagnostic potential. Furthermore, it would be interesting to combine findings across different –omics fields and explore the potentially complementary roles of genomic and proteomic biomarkers in comprising a disease-specific profile and enabling differential diagnosis, particularly given the likelihood that individual biomarkers could be shared with other psychiatric disorders such as BD and schizophrenia [194]. Exploring additional non-biological patient characteristics for their potential associations with MDD would also be of interest.

6.4. Conclusions

The work carried out in this thesis represents extensive investigations of proteomic biomarker candidates, sociodemographic factors and other patient characteristics to develop risk prediction models of MDD that could potentially be used to assist in the detection and diagnosis of MDD in clinical practice. Targeted MRM-MS was used to measure the abundances of a panel of peptides representing proteins which have been previously associated with psychiatric disorders in serum and DBS samples, and various non-biological patient characteristics were investigated as potential predictors of MDD along with proteomic biomarker candidates. By developing and adopting a multimodel approach, involving LASSO regression, repeated *K*-fold cross-validation, feature extraction and model averaging, parsimonious and reproducible prediction models were obtained. Several blood-based proteomic biomarkers, which mainly represented an immune/inflammatory profile, BMI, childhood trauma, as well as other features, were found to be associated with MDD. Although further investigations are needed, these associations indicate a dysregulation of the immune/inflammatory and neuroendocrine systems in the pathophysiology of depression. Moreover, these findings demonstrate that prediction models comprised of blood-based proteomic biomarkers, sociodemographic factors and clinical/personality features can identify and predict depression among clinically relevant populations of individuals presenting with subthreshold depression/low mood, despite the need for additional validation studies. The findings also demonstrate the utility of serum and DBS as sample materials for identifying blood-based biomarkers of MDD, and represent a substantial step forward in the development of an objective, non-invasive and cost-effective diagnostic test for MDD. Such a test has the potential to facilitate earlier and more accurate clinical diagnosis, particularly in primary care where time and resources are limited, and reduce the overall socioeconomic burden of MDD. Nevertheless, a considerable amount of further work remains to be done for a diagnostic test to be translated into routine clinical practice and for the pathophysiological mechanisms of MDD to be fully elucidated.

Appendix

Table A. 1. Protein peptides investigated using targeted MRM-MS in Chapters 3-5.

One hundred and forty-seven peptides representing 77 serum proteins were investigated in **Chapter 3**; One hundred and forty-six peptides representing 77 serum proteins were investigated in **Chapter 4**. One hundred and ninety-four peptides representing 115 DBS proteins were investigated in **Chapter 5**.

Protein	Protein name	UniProt accession number	Peptide sequence	Chapter 3 (serum)	Chapter 4 (serum)	Chapter 5 (DBS)
A1AG1	α-1-acid glycoprotein 1	P02763	SDVVYTDWK			✓
A1AG2	α-1-acid glycoprotein 2	P19652	EHVAHLLFLR			
A1AT	α-1-antitrypsin	P01009	LSITGTYDLK	✓	✓	✓
A1AT	α-1-antitrypsin	P01009	SPLFMGK	✓	✓	✓
A1AT	α-1-antitrypsin	P01009	SVLGQLGITK	✓	✓	✓
A1BG	α-1B-glycoprotein	P04217	ATWSGAVLAGR	✓	✓	✓
A1BG	α-1B-glycoprotein	P04217	CLAPLEGAR	✓	✓	
A1BG	α-1B-glycoprotein	P04217	GVTFLLR			✓
A1BG	α-1B-glycoprotein	P04217	LLELTGPK			✓
A1BG	α-1B-glycoprotein	P04217	SGLSTGWTQLSK	✓	✓	✓
A2AP	α-2-antiplasmin	P08697	DFLQSLK	✓	✓	✓
A2AP	α-2-antiplasmin	P08697	DSFHLDEQFTVPVEMMQAR	✓	✓	
A2AP	α-2-antiplasmin	P08697	FDPSLTQR	✓	✓	✓
A2AP	α-2-antiplasmin	P08697	LFGPDLK			✓
A2GL	Leucine-rich α-2-glycoprotein	P02750	VAAGAFQGLR			✓
A2MG	α-2-macroglobulin	P01023	AIGYLNTGYQR	✓	✓	✓
A2MG	α-2-macroglobulin	P01023	NEDSLVFVQTDK	✓	✓	✓
AACT	α-1-antichymotrypsin	P01011	ADLSGITGAR	✓	✓	✓
AACT	α-1-antichymotrypsin	P01011	EIGELYLPK	✓	✓	✓
AACT	α-1-antichymotrypsin	P01011	EQLSLLDR	✓	✓	✓

AL1A1	Retinal dehydrogenase 1	P00352	ILDLIESGK			✓
ALBU	Albumin	P02768	AAFTECCQAADK	✓	✓	✓
ALBU	Albumin	P02768	ETYGEMADCCAK	✓	✓	✓
ALBU	Albumin	P02768	QNCELFEQLGEYK	✓	✓	✓
ALBU	Albumin	P02768	YLYEIAR			✓
ALDOA	Fructose-bisphosphate aldolase A	P04075	ALANSLACQGK			✓
ALDOA	Fructose-bisphosphate aldolase A	P04075	QLLLTADDR			✓
AMBP	Protein AMBP	P02760	ETLLQDFR	✓	✓	✓
AMBP	Protein AMBP	P02760	TVAACNLPIVR	✓	✓	✓
ANGT	Angiotensinogen	P01019	ALQDQLVLVAAK	✓	✓	✓
ANGT	Angiotensinogen	P01019	FMQAVTGWK	✓	✓	
ANGT	Angiotensinogen	P01019	SLDFTELDVAAEK	✓	✓	✓
ANT3	Antithrombin-III	P01008	EVPLNTIIFMGR			✓
ANT3	Antithrombin-III	P01008	FATTFYQHLADSK			✓
ANT3	Antithrombin-III	P01008	FDTISEK	✓	✓	✓
ANT3	Antithrombin-III	P01008	LPGIVAEGR	✓	✓	✓
APOA1	Apolipoprotein A-I	P02647	ATEHLSTLSEK	✓	✓	✓
APOA1	Apolipoprotein A-I	P02647	EQLGPVTQEFWDNLEK	✓	✓	
APOA1	Apolipoprotein A-I	P02647	LLDNWDSVTSTFSK			✓
APOA1	Apolipoprotein A-I	P02647	VSFLSALEEYTK			✓
APOA2	Apolipoprotein A-II	P02652	EQLTPLIK			✓
APOA2	Apolipoprotein A-II	P02652	SPELQAEAK	✓	✓	✓
APOA4	Apolipoprotein A-IV	P06727	ALVQQMEQLR	✓	✓	✓
APOA4	Apolipoprotein A-IV	P06727	IDQNVEELK	✓	✓	✓
APOA4	Apolipoprotein A-IV	P06727	ISASAEELR	✓	✓	✓
APOA4	Apolipoprotein A-IV	P06727	LAPLAEDVR			✓
APOA4	Apolipoprotein A-IV	P06727	LLPHANEVSQK			✓
APOB	Apolipoprotein B-100	P04114	TGISPLALIK			✓
APOC1	Apolipoprotein C-I	P02654	EFGNTLEDK	✓	✓	✓
APOC1	Apolipoprotein C-I	P02654	EWFSETFQK	✓	✓	✓
APOC2	Apolipoprotein C-II	P02655	ESLSSYWESAK	✓	✓	
APOC2	Apolipoprotein C-II	P02655	TAAQNLYEK	✓	✓	✓

APOC2	Apolipoprotein C-II	P02655	TYLPAVDEK			✓
APOC3	Apolipoprotein C-III	P02656	DALSSVQESQVAQQAR	✓	✓	
APOC3	Apolipoprotein C-III	P02656	GWVTDGFSSLK	✓	✓	✓
APOC4	Apolipoprotein C-IV	P55056	AWFLESK	✓	✓	
APOD	Apolipoprotein D	P05090	NILTSNNIDVK			✓
APOD	Apolipoprotein D	P05090	VLNQELR	✓	✓	✓
APOE	Apolipoprotein E	P02649	AATVGSLAGQPLQER	✓	✓	✓
APOE	Apolipoprotein E	P02649	ALMDETMK	✓	✓	✓
APOE	Apolipoprotein E	P02649	LEEQAQQIR	✓	✓	
APOE	Apolipoprotein E	P02649	LGPLVEQGR	✓	✓	✓
APOE	Apolipoprotein E	P02649	SELEEQLTPVAEETR	✓	✓	✓
APOF	Apolipoprotein F	Q13790	SLPTEDCENEK	✓	✓	
APOH	β-2-glycoprotein 1	P02749	EHSSLAFWK	✓	✓	✓
APOH	β-2-glycoprotein 1	P02749	VSFFCK	✓	✓	
APOL1	Apolipoprotein L1	O14791	LNILNNNYK	✓	✓	✓
APOL1	Apolipoprotein L1	O14791	VNEPSILEMSR	✓	✓	✓
APOL1	Apolipoprotein L1	O14791	VTEPISAESGEQVER	✓	✓	
APOM	Apolipoprotein M	O95445	AFLLTPR	✓	✓	
APOM	Apolipoprotein M	O95445	FLLYNR			✓
APOM	Apolipoprotein M	O95445	SLTSCLDSK	✓	✓	✓
C1QB	Complement C1q subcomponent subunit B	P02746	GNLCVNLMR			✓
C1QC	Complement C1q subcomponent subunit C	P02747	FQSVFTVTR			✓
C1QC	Complement C1q subcomponent subunit C	P02747	TNQVNSGGVLLR	✓	✓	✓
C1R	Complement C1r	P00736	YTTEIIK	✓	✓	✓
C1RL	Complement C1r subcomponent-like protein	Q9NZP8	GSEAINAPGDNPAK	✓		

C1RL	Complement C1r subcomponent-like protein	Q9NZP8	VVVHPDYR			✓
C1S	Complement C1s subcomponent	P09871	LLEVPEGR	✓	✓	✓
C1S	Complement C1s subcomponent	P09871	TNFDNDIALVR	✓	✓	
C4BPA	C4b-binding protein α chain	P04003	EDVYVVGTVLR	✓	✓	✓
C4BPA	C4b-binding protein α chain	P04003	FSAICQGDGTWSPR	✓	✓	
C4BPA	C4b-binding protein α chain	P04003	GYILVGQAK			✓
C4BPA	C4b-binding protein α chain	P04003	YTCLPGYVR	✓	✓	✓
CAH1	Carbonic anhydrase 1	P00915	ADGLAVIGVLMK	✓	✓	✓
CAH2	Carbonic anhydrase 2	P00918	GGPLDGTYR			✓
CAH2	Carbonic anhydrase 2	P00918	SADFTNFDPR			✓
CAH2	Carbonic anhydrase 2	P00918	VVDVLDSIK			✓
CATA	Catalase	P04040	LNVITVGPR			✓
CBG	Corticosteroid-binding globulin	P08185	GTWTQPFDLASTR	✓	✓	✓
CBG	Corticosteroid-binding globulin	P08185	ITQDAQLK	✓	✓	
CBPB2	Carboxypeptidase B2	Q96IY4	DTGTYGFLLPER	✓	✓	
CBPB2	Carboxypeptidase B2	Q96IY4	YPLYVLK	✓	✓	
CD44	CD44 antigen	P16070	YGFIEGHVVIPR			✓
CD5L	CD5 antigen-like	O43866	EATLQDCPSGPWGK	✓	✓	
CD5L	CD5 antigen-like	O43866	IWLDNVR			✓
CD5L	CD5 antigen-like	O43866	LVGGLHR			✓
CERU	Ceruloplasmin	P00450	DIASGLIGPLIICK			✓
CERU	Ceruloplasmin	P00450	EVGPTNADPVCLAK	✓	✓	
CERU	Ceruloplasmin	P00450	EYTDASFTNR			✓

CERU	Ceruloplasmin	P00450	NNEGTYYSPNYNPQSR	✓	✓	
CFAB	Complement factor B	P00751	DISEVVTPR	✓	✓	✓
CFAB	Complement factor B	P00751	DLLYIGK	✓	✓	✓
CFAB	Complement factor B	P00751	EELLPAQDIK	✓	✓	✓
CFAB	Complement factor B	P00751	YGLVTYATYPK	✓	✓	✓
CFAH	Complement factor H	P08603	CFEGFGIDGPAIAK	✓	✓	
CFAH	Complement factor H	P08603	IDVHLVPDR			✓
CFAH	Complement factor H	P08603	NGFYPATR			✓
CFAI	Complement factor I	P05156	IVIEYVDR			✓
CHSP1	Calcium-regulated heat-stable protein 1	Q9Y2V2	LQAVEVVITHLAPGTK			
CLUS	Clusterin	P10909	ASSIIDELFQDR			✓
CLUS	Clusterin	P10909	ELDESLQVAER			✓
CLUS	Clusterin	P10909	FMETVAEK	✓	✓	✓
CLUS	Clusterin	P10909	IDSLLENDR	✓	✓	✓
CO2	Complement C2	P06681	HAIILLTDGK	✓	✓	
CO3	Complement C3	P01024	AGDFLEANYMNLQR	✓	✓	✓
CO3	Complement C3	P01024	GYTQQLAFR			✓
CO3	Complement C3	P01024	SGIPIVTSPYQIHFTK			
CO3	Complement C3	P01024	TGLQEVEVK		✓	✓
CO3	Complement C3	P01024	VYAYYNLEESCTR	✓	✓	
CO4A	Complement C4-A	P0C0L4	DFALLSLQVPLK	✓	✓	✓
CO4A	Complement C4-A	P0C0L4	ITQVLHFTK	✓	✓	
CO4A	Complement C4-A	P0C0L4	VGDTLNLNLR		✓	✓
CO4A	Complement C4-A	P0C0L4	VLSLAQEQVGGSPEK	✓	✓	✓
CO6	Complement component C6	P13671	SEYGAALAWEK	✓	✓	
CO6	Complement component C6	P13671	TLNICEVGTIR	✓	✓	
CO8A	Complement component C8 α chain	P07357	AMAVEDIISR			✓

CO8A	Complement component C8 α chain	P07357	HTSLGPLEAK			✓
CO8A	Complement component C8 α chain	P07357	MESLGITSR	✓	✓	
CO8B	Complement component C8 β chain	P07358	SGFSFGFK			✓
CO9	Complement component C9	P02748	LSPIYNLVPVK	✓	✓	✓
CO9	Complement component C9	P02748	VVEESELAR	✓	✓	✓
CXCL7	Platelet basic protein	P02775	NIQSLEVIGK			✓
DEMA	Dematin	Q08495	VTSNLGK			✓
ENOA	α-enolase	P06733	TIAPALVSK			✓
F13A	Coagulation factor XIII A chain	P00488	STVLTIPEIIIK			✓
FA12	Coagulation factor XII	P00748	CFEPQLLR	✓	✓	
FA12	Coagulation factor XII	P00748	VVGGLVALR	✓	✓	
FCN3	Ficolin-3	O75636	YGIDWASGR	✓	✓	
FETUA	α-2-HS-glycoprotein	P02765	FSVVYAK	✓	✓	
FETUA	α-2-HS-glycoprotein	P02765	HTLNQIDEVK	✓	✓	
FETUB	Fetuin-B	Q9UGM5	LVVLPFPK			✓
FIBA	Fibrinogen α chain	P02671	GSESGIFTNTK			✓
FIBB	Fibrinogen β chain	P02675	AHYGGFTVQNEANK			✓
FIBG	Fibrinogen γ chain	P02679	EGFGHLSPTGTTEFWLGNEK			✓
FIBG	Fibrinogen γ chain	P02679	IHLISTQSAIPYALR			
FINC	Fibronectin	P02751	SYTITGLQPGTDYK			✓
FINC	Fibronectin	P02751	YSFCTDHTVLVQTR	✓	✓	✓
GDIB	Rab GDP dissociation inhibitor β	P50395	DLGTESQIFISR			✓
GDIB	Rab GDP dissociation inhibitor β	P50395	FVSISDLLVPK			✓
GELS	Gelsolin	P06396	AGALNSNDAFVLK	✓	✓	✓
GELS	Gelsolin	P06396	SEDCFILDHGK	✓	✓	✓
H4	Histone H5	P62805	DAVTYTEHAK			✓

H4	Histone H4	P62805	VFLENVIR			✓
HABP2	Hyaluronan-binding protein 2	Q14520	VVLGDQDLK			✓
HBA	Hemoglobulin subunit α	P69905	FLASVSTVLTSK	✓	✓	
HBA	Hemoglobulin subunit α	P69905	MFLSFPTTK	✓	✓	
HBD	Hemoglobulin subunit δ	P02042	LLGNVLVCVLAR			
HBG1	Hemoglobulin subunit γ-1	P69891	MVTAVASALSSR	✓		✓
HEMO	Hemopexin	P02790	NFPSPVDAAFR	✓	✓	✓
HEMO	Hemopexin	P02790	VDGALCMEK	✓	✓	✓
HEP2	Heparin cofactor 2	P05546	FAFNLYR	✓	✓	✓
HEP2	Heparin cofactor 2	P05546	IAIDLFK	✓	✓	✓
HEP2	Heparin cofactor 2	P05546	TLEAQLTPR			✓
HINT1	Histidine triad nucleotide-binding protein 1	P49773	IIFEDDR			✓
HPT	Haptoglobin	P00738	DYAEVGR	✓	✓	✓
HPT	Haptoglobin	P00738	VGYVSGWGR	✓	✓	✓
HPT	Haptoglobin	P00738	VTSIQDWVQK	✓	✓	✓
HRG	Histidine-rich glycoprotein	P04196	ADLFYDVEALDLESPK	✓	✓	
HRG	Histidine-rich glycoprotein	P04196	DSPVLIDFFEDTER	✓	✓	
HRG	Histidine-rich glycoprotein	P04196	GGEGTGYFVDFSVR			✓
IC1	Plasma protease C1 inhibitor	P05155	FQPTLLTLPR	✓	✓	✓
IC1	Plasma protease C1 inhibitor	P05155	LLDSLPSDTR			✓
IC1	Plasma protease C1 inhibitor	P05155	TNLESILSYPK	✓		✓
IF4B	Eukaryotic translation initiation factor 4B	P23588	SILPTAPR			✓
IGHA1	Immunoglobulin heavy constant α 1	P01876	DASGVTFTWTPSSGK	✓	✓	
IGHA1	Immunoglobulin heavy constant α 1	P01876	TPLTATLSK	✓	✓	✓

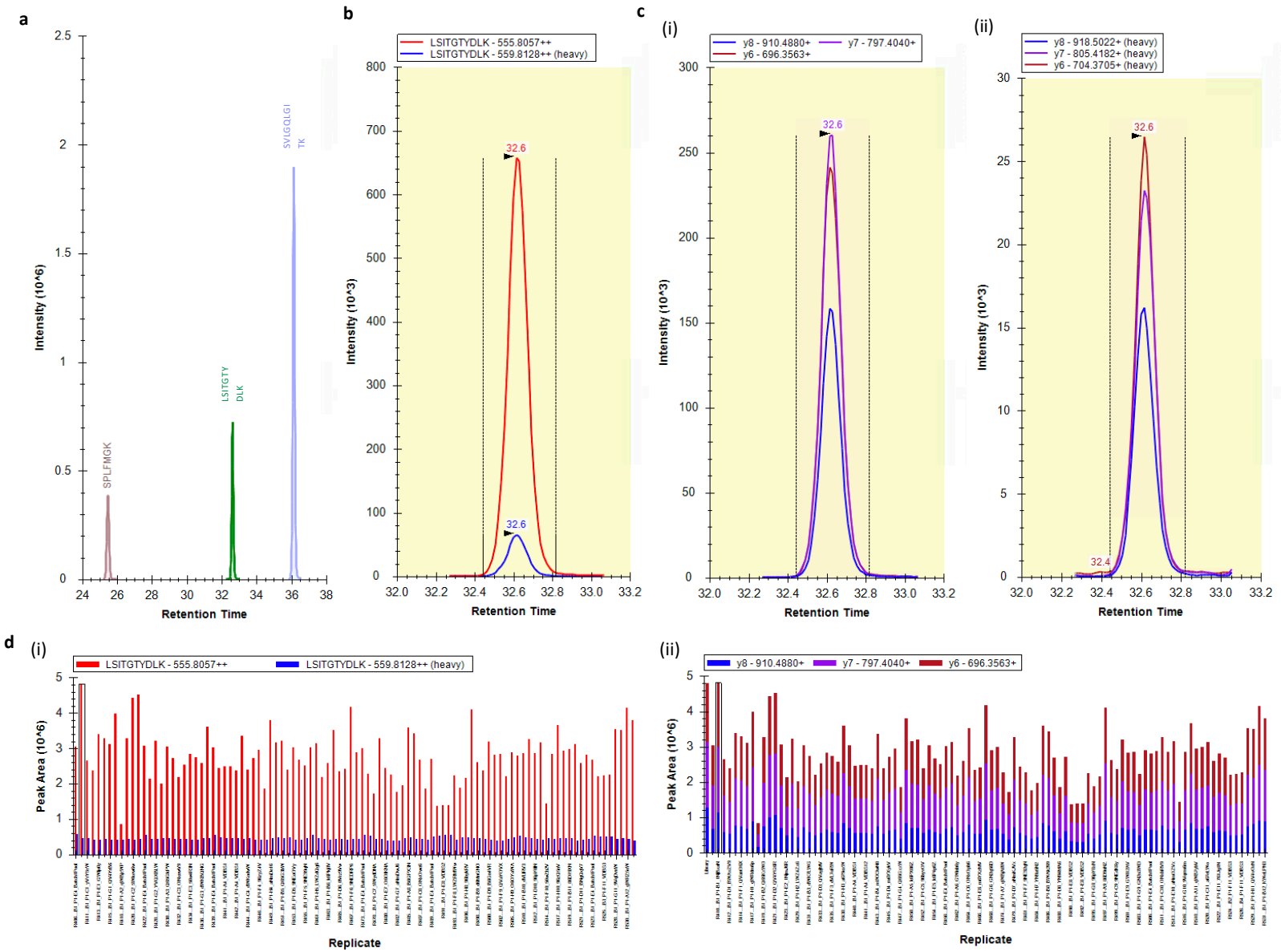
IGHA2	Immunoglobulin heavy constant α 2	P01877	DASGATFTWTPSSGK	✓	✓	
IGHG1	Immunoglobulin heavy constant γ 1	P01857	FNWYVDGVEVHNAK	✓	✓	✓
IGHG1	Immunoglobulin heavy constant γ 1	P01857	GPSVFPLAPSSK			✓
IGHG2	Immunoglobulin heavy constant γ 2	P01859	GLPAPIEK	✓	✓	
IGHG2	Immunoglobulin heavy constant γ 2	P01859	TTPPMLDSDGSFFLYSK	✓		
IGHG3	Immunoglobulin heavy constant γ 3	P01860	DTLMISR	✓	✓	✓
IGHG3	Immunoglobulin heavy constant γ 3	P01860	NQVSLTCLVK	✓	✓	✓
IGHM	Immunoglobulin heavy constant μ	P01871	GFPSVLR			✓
IGHM	Immunoglobulin heavy constant μ	P01871	QIQVSWLR	✓	✓	✓
IGHM	Immunoglobulin heavy constant μ	P01871	YAATSQVLLPSK	✓	✓	✓
ITIH1	Inter-α-trypsin inhibitor heavy chain H1	P19827	AAISGENAGLVR			✓
ITIH1	Inter-α-trypsin inhibitor heavy chain H1	P19827	GSLVQASEANLQAAQDFVR	✓	✓	
ITIH1	Inter-α-trypsin inhibitor heavy chain H1	P19827	LDAQASFLPK	✓	✓	
ITIH2	Inter-α-trypsin inhibitor heavy chain H2	P19823	FYNQVSTPLLR	✓	✓	
ITIH2	Inter-α-trypsin inhibitor heavy chain H2	P19823	IQPSGGTNINEALLR	✓	✓	
ITIH4	Inter-α-trypsin inhibitor heavy chain H4	Q14624	ETLFSVMPGLK	✓	✓	
ITIH4	Inter-α-trypsin inhibitor heavy chain H4	Q14624	GPDVLTATVSGK	✓	✓	✓
ITIH4	Inter-α-trypsin inhibitor heavy chain H4	Q14624	ILDDLSPR			✓

KAD1	Adenylate kinase isoenzyme 1	P00568	IIFVVGGPGSGK			✓
KLKB1	Plasma kallikrein	P03952	LSMDGSPTR	✓	✓	
KNG1	Kininogen-1	P01042	DFVQPPTK	✓		
KNG1	Kininogen-1	P01042	DIPTNSPELEETLTHTITK	✓	✓	✓
KNG1	Kininogen-1	P01042	TVGSDTFYSFK			✓
KNG1	Kininogen-1	P01042	YFIDFVAR			✓
KPYR	KPYR_HUMAN	P30613	GDLGIEIPAEK			✓
LEG3	Galectin-3	P17931	IALDFQR			✓
LUM	Lumican	P51884	SLEDLQLTHNK	✓	✓	✓
MUCB	Immunoglobulin heavy constant μ	P01871	GQPLSPEK			✓
NDKA	Nucleoside diphosphate kinase A	P15531	DRPFFAGLVK			✓
PARK7	Parkinson disease protein 7	Q99497	ALVILAK			✓
PARK7	Parkinson disease protein 7	Q99497	DGLILTSR			✓
PEBP1	Phosphatidylethanolamine-binding protein 1	P30086	LYEQLSGK			✓
PEDF	Pigment epithelium-derived factor	P36955	DTDTGALLFIGK	✓	✓	
PEDF	Pigment epithelium-derived factor	P36955	ELLDTVTAPQK	✓	✓	✓
PEDF	Pigment epithelium-derived factor	P36955	LQSLFDSPDFSK	✓	✓	
PEDF	Pigment epithelium-derived factor	P36955	TVQAVLTVPK	✓	✓	✓
PERM	Myeloperoxidase	P05164	IANVFTNAFR			✓
PGRP2	N-acetylmuramoyl-L-alanine amidase	Q96PD5	GCPDVQASLPDAK	✓		
PGRP2	N-acetylmuramoyl-L-alanine amidase	Q96PD5	TFTLLDPK	✓	✓	

PHLD	Phosphatidylinositol-glycan-specific phospholipase D	P80108	NQVVIAAGR	✓	✓	
PLMN	Plasminogen	P00747	FVTWIEGVMR	✓	✓	✓
PLMN	Plasminogen	P00747	HSIFTPETNPR			
PMGE	Bisphosphoglycerate mutase	P07738	HYGALIGLNR			✓
PNPH	Purine nucleoside phosphorylase	P00491	FEVGDIMLIR			✓
PNPH	Purine nucleoside phosphorylase	P00491	VFGFSLITNK			✓
PON1	Serum paraoxonase/arylesterase 1	P27169	LLIGTVFHK			✓
PPAC	Low molecular weight phosphotyrosine protein phosphatase	P24666	IELLGSYDPQK			✓
PPIA	Peptidyl-prolyl cis-trans isomerase A	P62937	FEDENFILK			✓
PRDX1	Peroxiredoxin-1	Q06830	ADEGISFR			✓
PRDX2	Peroxiredoxin-2	P32119	SVDEALR			✓
PRDX2	Peroxiredoxin-2	P32119	TDEGIAYR			✓
PRDX6	Peroxiredoxin-6	P30041	LSILYPATTGR			✓
PROF1	Profilin-1	P07737	TLVLLMGK			✓
PRS6A	26S proteasome regulatory subunit 6A	P17980	DAFALAK			✓
PRS8	26S proteasome regulatory subunit 8	P62195	FIGEGAR			✓
PSA2	Proteasome subunit α type-2	P25787	AANGVVLATEK			✓
PSA6	Proteasome subunit α type-6	P60900	HITIFSPEGR			✓
PSB7	Proteasome subunit β type-7	Q99436	GTTAVLTEK			✓
RANG	Ran-specific GTPase-activating protein	P43487	FLNAENAQK			✓

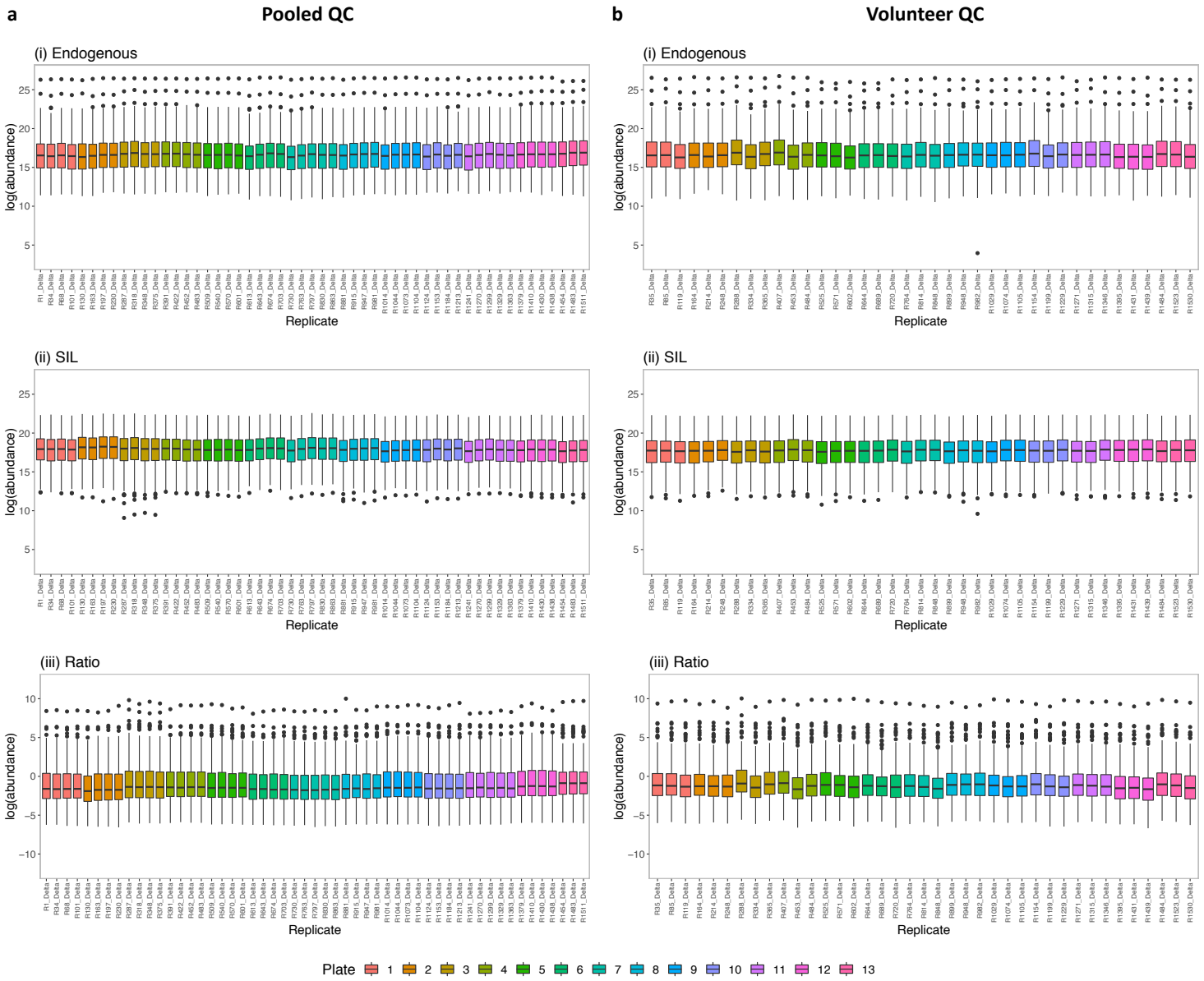
RET4	Retinol-binding protein 4	P02753	QEELCLAR	✓	✓	
RET4	Retinol-binding protein 4	P02753	YWGVASFLQK	✓		
SAA4	Serum amyloid A-4 protein	P35542	EALQGVGDMGR			✓
SAMP	Serum amyloid P-component	P02743	IVLGQEQDSYGGK	✓	✓	✓
SH3L3	SH3 domain-binding glutamic acid-rich-like protein 3	Q9H299	VYSTSVTGSR			✓
SHBG	Sex hormone-binding globulin	P04278	IALGGLLFPASNLR	✓	✓	
SHBG	Sex Hormone-binding globulin	P04278	LPLVPALDGCLR		✓	
TBA4A	Tubulin α-4A chain	P68366	EIIDPVLDR			✓
TETN	Tetranectin	P05452	EQQALQTVCLK	✓	✓	
THBG	Thyroxine-binding globulin	P05543	NALALFVLPK			✓
THIO	Thioredoxin	P10599	VGEFSGANK			✓
THRB	Prothrombin	P00734	ELLESYIDGR	✓	✓	✓
THRB	Prothrombin	P00734	SGIECQLWR	✓	✓	
TPIS	Triosephosphate isomerase	P60174	FFVGGNWK			✓
TRFE	Serotransferrin	P02787	EGYYGYTGAFR	✓	✓	✓
TRFE	Serotransferrin	P02787	MYLGYEYVTAIR			✓
TSP1	Thrombospondin-1	P07996	GTLLALER			✓
TTHY	Transthyretin	P02766	AADDTWEPFASGK	✓	✓	
TTHY	Transthyretin	P02766	VLDAVR	✓	✓	
UB2L3	Ubiquitin-conjugating enzyme E2 L3	P68036	IYHPNIDEK			✓
VTDB	Vitamin D-binding protein	P02774	HLSLLTTLSNR		✓	✓
VTDB	Vitamin D-binding protein	P02774	THLPEVFLSK		✓	
VTDB	Vitamin D-binding protein	P02774	VLEPTLK		✓	✓

VTNC	Vitronectin	P04004	DVWGIEGPIDAAFTR	✓	✓	✓
VTNC	Vitronectin	P04004	DWHGVPGQVDAAMAGR	✓	✓	✓


a
Intensity (10^6)
Retention Time
SPLFMGK
LSITGTY
DLK
SVLGQLGI
TK
b
LSITGTYDLK - 555.8057++
LSITGTYDLK - 559.8128++ (heavy)
Intensity (10^3)
Retention Time
32.6
32.6
c
(i)
y8 - 910.4880+
y7 - 797.4040+
y6 - 696.3563+
Intensity (10^3)
Retention Time
32.6
(ii)
y8 - 918.5022+ (heavy)
y7 - 805.4182+ (heavy)
y6 - 704.3705+ (heavy)
Intensity (10^3)
Retention Time
32.6
32.4
d
(i)
LSITGTYDLK - 555.8057++
LSITGTYDLK - 559.8128++ (heavy)
Peak Area (10^6)
Replicate
(ii)
y8 - 910.4880+
y7 - 797.4040+
y6 - 696.3563+
Peak Area (10^6)
Replicate

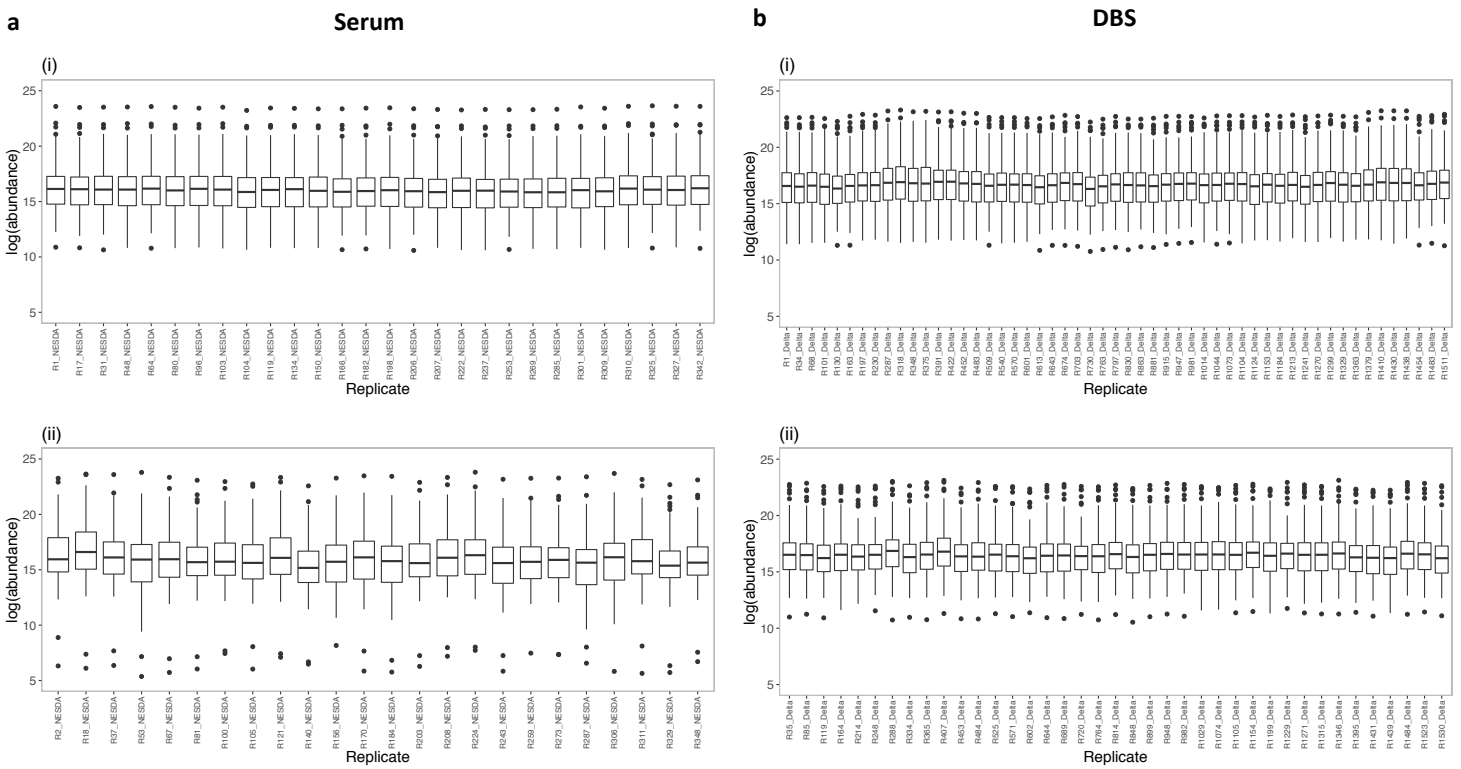
Appendix Figure 1. Skyline MS/MS results from targeted MRM-based protein peptide quantification using QQQ mass spectrometer.

(a) Extracted ion chromatogram for three peptides SPLFMGK, LSITGTYDLK and SVLGQLGITK targeted for protein α-1-antitrypsin (A1AT) in a digested DBS sample. (b) Precursor (MS1) ion chromatogram for peptide LSITGTYDLK (monitored at retention time 32.6 minutes). The endogenous peptide is shown in red and the SIL (heavy) peptide is shown in blue. Peak area was determined for each ion as the total integrated area within peak boundaries (vertical dotted lines), minus background area. (c) Fragment (MS2) ion chromatogram for y6, y7 and y8 transitions of the (i) endogenous and (ii) SIL peptide LSITGTYDLK (monitored at retention time 32.6 minutes). The y7 (purple) fragment ion served as a quantifier transition and was used for quantitation, whereas the y6 (red) and y8 (blue) fragment ions served as qualifier transitions and were used to ensure the correct selection of the endogenous and SIL peptides. Peak area was determined for each ion as the total integrated area within peak boundaries (vertical dotted lines), minus background area. (d) Peak area determined from: (i) peptide precursors and (iI) fragments across multiple digested DBS samples. Abbreviations: MRM (multiple reaction monitoring); MS (mass spectrometry); QQQ (triple quadrupole); SIL (stable isotope-labelled).


a
Pooled QC
b
Volunteer QC
(i) Endogenous
(ii) SIL
(iii) Ratio
log(abundance)
Replicate
Plate 1 2 3 4 5 6 7 8 9 10 11 12 13

Appendix Figure 2. Peptide-transition abundances of QC samples.

QC samples were used to assess the technical variation associated with instrument performance and sample preparation. (a) The variation in instrument performance was assessed using a pooled QC sample, which was prepared by pooling together the digested clinical samples and injected once a day along with the clinical samples for the duration of the entire study. (b) The variation in sample preparation was assessed using QC samples obtained from a healthy volunteer, which were prepared following the same protocol as the clinical samples and distributed across the experimental plates. Peptide-transition abundance (peak area) data are shown for 194 peptides representing 115 DBS proteins (measured in **Chapter 5**). Boxplots show the log_2-transformed abundance values of the peptide-transitions: (i) endogenous, (ii) SIL, and (iii) ratio. Boxplots are coloured by plate. Abbreviations: DBS (dried blood spot); QC (quality control).


a
Serum
b
DBS
(i)
(ii)
log(abundance)
Replicate

Appendix Figure 3. Endogenous peptide-transition abundances in serum and DBS QC samples. Peptide-transition abundance (peak area) data are shown for 89 endogenous peptides representing 51 proteins that were measured in both (a) serum and (b) DBS samples. QC samples were used to assess the technical variation associated with instrument performance and sample preparation. (i) The variation in instrument performance was assessed using a pooled QC sample, which was prepared by pooling together the digested clinical serum or DBS samples and injected once a day along with the clinical samples for the duration of the entire study. (ii) The variation in sample preparation was assessed using commercial serum samples (Human Sera S7023, Sigma Aldrich) or DBS samples obtained from a healthy volunteer, which were prepared following the same protocol as the clinical samples and distributed across the experimental plates. Boxplots show the $\log_2$-transformed abundance values of the endogenous peptide-transitions. Abbreviations: DBS (dried blood spot); QC (quality control).

References

[1] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders, fifth edition*. American Psychiatric Association, 2013.

[2] World Health Organization, "Depression and other common mental disorders: global health estimates," *World Heal. Organ.*, pp. 1–24, 2017.

[3] R. C. Kessler *et al.*, "The Epidemiology of Major Depressive Disorder: Results From the National Comorbidity Survey Replication (NCS-R)," *JAMA*, vol. 289, no. 23, p. 3095, Jun. 2003.

[4] R. C. Kessler, "Lifetime and 12-Month Prevalence of DSM-III-R Psychiatric Disorders in the United States," *Arch. Gen. Psychiatry*, vol. 51, no. 1, p. 8, Jan. 1994.

[5] E. Bromet *et al.*, "Cross-national epidemiology of DSM-IV major depressive episode," *BMC Med.*, vol. 9, p. 90, Jul. 2011.

[6] World Health Organization, "Mental Health and Development: Targeting People with Mental Health Conditions as a Vulnerable Group," *World Heal. Organ.*, 2013.

[7] B. W. Penninx, Y. Milaneschi, F. Lamers, and N. Vogelzangs, "Understanding the somatic consequences of depression: Biological mechanisms and the role of depression symptom profile," *BMC Med.*, vol. 11, no. 1, 2013.

[8] P. Cuijpers, N. Vogelzangs, J. Twisk, A. Kleiboer, J. Li, and B. W. Penninx, "Comprehensive meta-analysis of excess mortality in depression in the general community versus patients with specific illnesses," *Am. J. Psychiatry*, vol. 171, no. 4, pp. 453–462, 2014.

[9] E. R. Walker, R. E. McGee, and B. G. Druss, "Mortality in mental disorders and global disease burden implications a systematic review and meta-analysis," *JAMA Psychiatry*, vol. 72, no. 4, pp. 334–341, 2015.

[10] E. Chesney, G. M. Goodwin, and S. Fazel, "Risks of all-cause and suicide mortality in mental disorders: A meta-review," *World Psychiatry*, vol. 13, no. 2, pp. 153–160, Jun. 2014.

[11] World Health Organization, "Mental Health and Development: Targeting People with Mental Health Conditions as a Vulnerable Group," *World Heal. Organ.*, 2010.

[12] J. Olesen, A. Gustavsson, M. Svensson, H. U. Wittchen, and B. Jönsson, "The economic cost of brain disorders in Europe," *Eur. J. Neurol.*, vol. 19, no. 1, pp. 155–162, 2012.

[13] K. Demyttenaere *et al.*, "Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys," *J. Am. Med. Assoc.*, vol. 291, no. 21, pp. 2581–2590, Jun. 2004.

[14] M. K. Chan *et al.*, "Applications of blood-based protein biomarker strategies in the study of psychiatric disorders," *Prog. Neurobiol.*, vol. 122, pp. 45–72, 2014.

[15] M. K. Chan, J. D. Cooper, and S. Bahn, "Commercialisation of Biomarker Tests for Mental Illnesses: Advances and Obstacles," *Trends Biotechnol.*, vol. 33, pp. 712–723, 2015.

[16] World Health Organization, "International Classification of Diseases - 11th Revision," *World Heal. Organ.*, vol. 11, 2018.

[17] A. J. Mitchell, A. Vaze, and S. Rao, "Clinical diagnosis of depression in primary care: a meta-analysis," *Lancet*, vol. 374, no. 9690, pp. 609–619, 2009.

[18] R. M. A. Hirschfeld *et al.*, "The National Depressive and Manic-Depressive Association consensus statement on the undertreatment of depression," in *Journal of the American Medical Association*, 1997, vol. 277, no. 4, pp. 333–340.

[19] P. S. Wang *et al.*, "Worldwide use of mental health services for anxiety, mood, and substance disorders: Results from 17 countries in the WHO world mental health surveys," *Lancet*, vol. 370, no. 9590, pp. 841–850, 2007.

[20] A. Tylee and P. Gandhi, "The importance of somatic symptoms in depression in primary care," *Prim. Care Companion J. Clin. Psychiatry*, vol. 7, no. 4, pp. 167–176, Aug. 2005.

[21] L. J. Kirmayer, J. M. Robbins, M. Dworkind, and M. J. Yaffe, "Somatization and the recognition of depression and anxiety in primary care," *Am. J. Psychiatry*, vol. 150, no. 5, pp. 734–741, Apr. 1993.

[22] C. Dowrick and A. Frances, "Medicalising unhappiness: new classification of depression risks more patients being put on drug treatment from which they will not benefit," *BMJ*, vol. 347, 2013.

[23] S. K. Davidson, M. G. Harris, C. F. Dowrick, C. A. Wachtler, J. Pirkis, and J. M. Gunn, "Mental health interventions and future major depression among primary care patients with subthreshold depression," *J. Affect. Disord.*, vol. 177, pp. 65–73, May 2015.

[24] R. Mojtabai, "Clinician-Identified Depression in Community Settings: Concordance with Structured-Interview Diagnoses," *Psychother. Psychosom.*, vol. 82, no. 3, pp. 161–169, Apr. 2013.

[25] R. C. Kessler and B. B. Üstün, "The World Mental Health (WMH) Survey Initiative version of the World Health Organization (WHO) Composite International Diagnostic

Interview (CIDI)," *Int. J. Methods Psychiatr. Res.*, vol. 13, no. 2, pp. 93–117, 2004.

[26] J. M. Haro *et al.*, "Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health Surveys," *Int. J. Methods Psychiatr. Res.*, vol. 15, no. 4, pp. 167–180, 2006.

[27] R. C. Kessler *et al.*, "Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMH-CIDI)," *Int. J. Methods Psychiatr. Res.*, vol. 13, no. 2, pp. 122–139, 2004.

[28] H. U. Wittchen, "Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review," *J. Psychiatr. Res.*, vol. 28, no. 1, pp. 57–84, 1994.

[29] R. C. Kessler and E. J. Bromet, "The Epidemiology of Depression Across Cultures," *Annu. Rev. Public Health*, vol. 34, no. 1, pp. 119–138, 2013.

[30] S. Seedat *et al.*, "Cross-national associations between gender and mental disorders in the World Health Organization World Mental Health Surveys," *Arch. Gen. Psychiatry*, vol. 66, no. 7, pp. 785–795, Jul. 2009.

[31] J. Spijker, R. Graaf, R. Bijl, and A. Beekman, "Duration of MDD episodes in the general population," *Br. J. Psychiatry*, vol. 181, no. Cidi, pp. 208–213, 2002.

[32] R. C. Kessler, W. T. Chiu, O. Demler, K. R. Merikangas, and E. E. Walters, "Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication," *Arch. Gen. Psychiatry*, vol. 62, no. 6, pp. 617–27, Jun. 2005.

[33] T. K. Melartin, H. J. Rytsälä, U. S. Leskelä, P. S. Lestelä-Mielonen, T. P. Sokero, and E. T. Isometsä, "Current comorbidity of psychiatric disorders among DSM-IV major depressive disorder patients in psychiatric care in the Vantaa Depression Study," *J. Clin. Psychiatry*, vol. 63, no. 2, pp. 126–134, 2002.

[34] K. Gao *et al.*, "Should an assessment of Axis i comorbidity be included in the initial diagnostic assessment of mood disorders? Role of QIDS-16-SR total score in predicting number of Axis i comorbidity," *J. Affect. Disord.*, vol. 148, no. 2–3, pp. 256–264, Jun. 2013.

[35] B. W. J. H. Penninx *et al.*, "Two-year course of depressive and anxiety disorders: Results from the Netherlands Study of Depression and Anxiety (NESDA)," *J. Affect. Disord.*, vol. 133, no. 1–2, pp. 76–85, Sep. 2011.

[36] J. G. F. M. Hovens, E. J. Giltay, J. E. Wiersma, P. Spinhoven, B. W. J. H. Penninx, and

F. G. Zitman, “Impact of childhood life events and trauma on the course of depressive and anxiety disorders,” *Acta Psychiatr. Scand.*, vol. 126, no. 3, pp. 198–207, 2012.

[37] D. J. Kupfer, “Long-term treatment of depression,” *J. Clin. Psychiatry*, vol. 52, no. 5 SUPPL., pp. 28–34, 1991.

[38] D. E. Giles, R. B. Jarrett, M. M. Biggs, D. S. Guzick, and A. J. Rush, “Clinical predictors of recurrence in depression,” *Am. J. Psychiatry*, vol. 146, no. 6, pp. 764–767, 1989.

[39] FDA, “Depression: Medicines To Help You,” *FDA Consumer Information*, 2009. [Online]. Available: http://www.fda.gov/downloads/ForConsumers/ByAudience/ForWomen/FreePublications/UCM182083.pdf. [Accessed: 30-Oct-2019].

[40] A. F. Leuchter, I. A. Cook, A. M. Hunter, and A. S. Korb, “A new paradigm for the prediction of antidepressant treatment response,” *Dialogues in Clinical Neuroscience*, vol. 11, no. 4. pp. 435–446, 2009.

[41] G. I. Papakostas and M. Fava, “Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD,” *Eur. Neuropsychopharmacol.*, vol. 19, no. 1, pp. 34–40, Jan. 2009.

[42] D. B. Miller and J. P. O’Callaghan, “Personalized medicine in major depressive disorder - Opportunities and pitfalls,” *Metabolism.*, vol. 62, no. SUPPL.1, pp. 34–39, 2013.

[43] B. N. Gaynes, D. Warden, M. H. Trivedi, S. R. Wisniewski, M. Fava, and A. J. Rush, “What did STAR*D teach us? Results from a large-scale, practical, clinical trial for patients with depression,” *Psychiatr. Serv.*, vol. 60, no. 11, pp. 1439–1445, 2009.

[44] G. I. Keitner, C. E. Ryan, I. W. Miller, and W. H. Norman, “Recovery and major depression: Factors associated with twelve-month outcome,” *Am. J. Psychiatry*, vol. 149, no. 1, pp. 93–99, 1992.

[45] C. Kraus, B. Kadriu, R. Lanzenberger, C. A. Zarate, and S. Kasper, “Prognosis and improved outcomes in major depression: a review,” *Transl. Psychiatry*, vol. 9, no. 1, 2019.

[46] M. Riedel *et al.*, “Clinical predictors of response and remission in inpatients with depressive syndromes,” *J. Affect. Disord.*, vol. 133, no. 1–2, pp. 137–149, Sep. 2011.

[47] National Collaborating Centre for Mental Health, “Depression: The NICE guideline on the treatment and management of depression in adults (updated edition),” 2010.

[48] C. Otte *et al.*, “Major depressive disorder,” *Nat. Rev. Dis. Prim.*, vol. 2, Sep. 2016.

[49] P. Cuijpers, M. Sijbrandij, S. L. Koole, G. Andersson, A. T. Beekman, and C. F. Reynolds, “Adding psychotherapy to antidepressant medication in depression and

anxiety disorders: A meta-analysis," *World Psychiatry*, vol. 13, no. 1, pp. 56–67, 2014.

[50] S. Pampallona, P. Bollini, G. Tibaldi, B. Kupelnick, and C. Munizza, "Combined pharmacotherapy and psychological treatment for depression: A systematic review," *Arch. Gen. Psychiatry*, vol. 61, no. 7, pp. 714–719, Jul. 2004.

[51] A. J. Gelenberg, "A review of the current guidelines for depression treatment," *J. Clin. Psychiatry*, vol. 71, no. 7, 2010.

[52] National Institute for Health and Care Excellence, "Depression in adults: recognition and management," 2019.

[53] T. B. Karasu, A. Gelenberg, A. Merriam, and P. Wang, "Treatment of Patients With Major Depressive Disorder, third edition," *Am. Psychiatr. Assoc.*, no. April, pp. 1–78, 2000.

[54] M. Zimmerman, I. Chelminski, J. B. McGlinchey, and D. Young, "Diagnosing major depressive disorder X: Can the utility of the DSM-IV symptom criteria be improved?," *J. Nerv. Ment. Dis.*, vol. 194, no. 12, pp. 893–897, Dec. 2006.

[55] L. L. Judd, H. S. Akiskal, and M. P. Paulus, "The role and clinical significance of subsyndromal depressive symptoms (SSD) in unipolar major depressive disorder," *J. Affect. Disord.*, vol. 45, no. 1–2, p. 27, 1997.

[56] N. Sadek and J. Bona, "Subsyndromal symptomatic depression: A new concept," *Depress. Anxiety*, vol. 12, no. 1, pp. 30–39, 2000.

[57] J. L. Ayuso-Mateos, R. Nuevo, E. Verdes, N. Naidoo, and S. Chatterji, "From depressive symptoms to depressive disorders: The relevance of thresholds," *Br. J. Psychiatry*, vol. 196, no. 5, pp. 365–371, 2010.

[58] M. R. Rodríguez, R. Nuevo, S. Chatterji, and J. L. Ayuso-Mateos, "Definitions and factors associated with subthreshold depressive conditions: A systematic review," *BMC Psychiatry*, vol. 12, 2012.

[59] L. L. Judd, M. H. Rapaport, M. P. Paulus, and J. L. Brown, "Subsyndromal symptomatic depression: A new mood disorder?," *J. Clin. Psychiatry*, vol. 55, no. 4 SUPPL., pp. 18–28, 1994.

[60] A. F. B. da Silva Lima and M. P. de Almeida Fleck, "Subsyndromal depression: An impact on quality of life?," *J. Affect. Disord.*, vol. 100, no. 1–3, pp. 163–169, 2007.

[61] M. Zimmerman, J. H. Martinez, K. Dalrymple, I. Chelminski, and D. Young, "'Subthreshold' Depression: Is the distinction between adjustment disorder with depressed mood and adjustment disorder with mixed anxious and depressed mood valid?," *J. Clin. Psychiatry*, vol. 25, no. 4, pp. 257–265, May 2013.

[62] M. Tuithof, M. ten Have, S. van Dorsselaer, M. Kleinjan, A. Beekman, and R. de Graaf, "Course of subthreshold depression into a depressive disorder and its risk factors," *J. Affect. Disord.*, vol. 241, pp. 206–215, Dec. 2018.

[63] M. Vuorilehto, T. Melartin, and E. T. Isometsä, "Depressive disorders in primary care: Recurrent, chronic, and co-morbid," *Psychol. Med.*, vol. 35, no. 5, pp. 673–682, May 2005.

[64] P. Rucci *et al.*, "Subthreshold psychiatric disorders in primary care: Prevalence and associated characteristics," *J. Affect. Disord.*, vol. 76, no. 1–3, pp. 171–181, 2003.

[65] E. J. Regeer, L. Krabbendam, R. De Graaf, M. Ten Have, W. A. Nolen, and J. Van Os, "A prospective study of the transition rates of subthreshold (hypo)mania and depression in the general population," *Psychol. Med.*, vol. 36, no. 5, pp. 619–627, May 2006.

[66] C. Gómez-Restrepo, A. Bohórquez, D. Pinto Masis, J. F. A. Gil Laverde, M. Rondón Sepúlveda, and N. Díaz-Granados, "The prevalence of and factors associated with depression in Colombia," *Rev. Panam. Salud Pública*, vol. 16, no. 6, pp. 378–386, Dec. 2004.

[67] P. Cuijpers and F. Smit, "Subthreshold depression as a risk indicator for major depressive disorder: A systematic review of prospective studies," *Acta Psychiatr. Scand.*, vol. 109, no. 5, pp. 325–331, 2004.

[68] R. D. Goldney, L. J. Fisher, E. Dal Grande, and A. W. Taylor, "Subsyndromal depression: Prevalence, use of health services and quality of life in an Australian population," *Soc. Psychiatry Psychiatr. Epidemiol.*, vol. 39, no. 4, pp. 293–298, Apr. 2004.

[69] C. D. Sherbourne, K. B. Wells, R. D. Hays, W. Rogers, M. A. Burnam, and L. L. Judd, "Subthreshold depression and depressive disorder: Clinical characteristics of general medical and mental health specialty outpatients," *Am. J. Psychiatry*, vol. 151, no. 12, pp. 1777–1784, Apr. 1994.

[70] Y. Forsell, "A three-year follow-up of major depression, dysthymia, minor depression and subsyndromal depression: Results from a population-based study," *Depress. Anxiety*, vol. 24, no. 1, pp. 62–65, Jan. 2007.

[71] N. Markkula *et al.*, "Predictors of new-onset depressive disorders – Results from the longitudinal Finnish Health 2011 Study," *J. Affect. Disord.*, vol. 208, pp. 255–264, Jan. 2017.

[72] R. H. Pietrzak, J. Kinley, T. O. Afifi, M. W. Enns, J. Fawcett, and J. Sareen, "Subsyndromal depression in the United States: Prevalence, course, and risk for incident

psychiatric outcomes," *Psychol. Med.*, vol. 43, no. 7, pp. 1401–1414, 2013.

[73] J. Fogel, W. W. Eaton, and D. E. Ford, "Minor depression as a predictor of the first onset of major depressive disorder over a 15-year follow-up," *Acta Psychiatr. Scand.*, vol. 113, no. 1, pp. 36–43, Jan. 2006.

[74] S. A. Shankman, P. M. Lewinsohn, D. N. Klein, J. W. Small, J. R. Seeley, and S. E. Altman, "Subthreshold conditions as precursors for full syndrome disorders: A 15-year longitudinal study of multiple diagnostic classes," *J. Child Psychol. Psychiatry Allied Discip.*, vol. 50, no. 12, pp. 1485–1494, 2009.

[75] E. Horwath, J. Johnson, G. L. Klerman, and M. M. Weissman, "Depressive Symptoms as Relative and Attributable Risk Factors for First-Onset Major Depression," *Arch. Gen. Psychiatry*, vol. 49, no. 10, pp. 817–823, 1992.

[76] Y. Y. Lee *et al.*, "The risk of developing major depression among individuals with subthreshold depression: a systematic review and meta-analysis of longitudinal cohort studies," *Psychol. Med.*, vol. 1, no. 11, 2018.

[77] P. Cuijpers, R. De Graaf, and S. Van Dorsselaer, "Minor depression: Risk profiles, functional disability, health care use and risk of developing major depression," *J. Affect. Disord.*, vol. 79, no. 1–3, pp. 71–79, Apr. 2004.

[78] K. van Zoonen *et al.*, "Preventing the onset of major depressive disorder: A meta-analytic review of psychological interventions," *Int. J. Epidemiol.*, vol. 43, no. 2, pp. 318–329, 2014.

[79] E. A. Stockings *et al.*, "Preventing depression and anxiety in young people: A review of the joint efficacy of universal, selective and indicated prevention," *Psychol. Med.*, vol. 46, no. 1, pp. 11–26, 2016.

[80] C. Buntrock *et al.*, "Effect of a Web-Based Guided Self-help Intervention for Prevention of Major Depression in Adults With Subthreshold Depression," *JAMA*, vol. 315, no. 17, p. 1854, 2016.

[81] D. Chisholm, K. Sanderson, J. L. Ayuso-Mateos, and S. Saxena, "Reducing the global burden of depression: Population-level analysis of intervention cost-effectiveness in 14 world regions," *Br. J. Psychiatry*, vol. 184, no. MAY, pp. 393–403, May 2004.

[82] B. G. Druss *et al.*, "Understanding mental health treatment in persons without mental diagnoses: Results from the National Comorbidity Survey replication," *Arch. Gen. Psychiatry*, vol. 64, no. 10, pp. 1196–1203, 2007.

[83] J. Angst and K. Merikangas, "The depressive spectrum in the community: Diagnostic classification and course," *Biol. Psychiatry*, vol. 42, pp. 220S-221S, 1997.

[84] G. Andrews, T. Brugha, M. E. Thase, F. F. Duffy, P. Rucci, and T. Slade, "Dimensionality and the category of major depressive episode," *Int. J. Methods Psychiatr. Res.*, vol. 16, no. SUPPL. 1, pp. S41–S51, Jun. 2007.

[85] D. Goldberg, "Plato versus Aristotle: Categorical and dimensional models for common mental disorders," *Compr. Psychiatry*, vol. 41, pp. 8–13, 2000.

[86] J. Angst, R. Sellar, and K. R. Merikangas, "Depressive spectrum diagnoses," *Compr. Psychiatry*, vol. 41, no. 2, pp. 39–47, 2000.

[87] A. Solomon, D. A. F. Haaga, and B. A. Arnow, "Is clinical depression distinct from subthreshold depressive symptoms? A review of the continuity issue in depression research," *J. Nerv. Ment. Dis.*, vol. 189, no. 8, pp. 498–506, 2001.

[88] S. D. Østergaard, S. O. W. Jensen, and P. Bech, "The heterogeneity of the depressive syndrome: When numbers get serious," *Acta Psychiatr. Scand.*, vol. 124, no. 6, pp. 495–496, Dec. 2011.

[89] E. I. Fried, R. M. Nesse, K. Zivin, C. Guille, and S. Sen, "Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors," *Psychol. Med.*, no. December, pp. 1–10, 2014.

[90] V. E. Cosgrove and T. Suppes, "Informing DSM-5: biological boundaries between bipolar I disorder, schizoaffective disorder, and schizophrenia," *BMC Med.*, vol. 11, no. 1, p. 127, May 2013.

[91] A. Jablensky, "The disease entity in psychiatry: fact or fiction?," *Epidemiol. Psychiatr. Sci.*, vol. 21, no. 3, pp. 255–264, 2012.

[92] D. Regier, "Merging categorical and dimensional diagnoses of mental disorders," *Epidemiol. Psychiatr. Sci.*, vol. 21, no. 3, pp. 267–269, Sep. 2012.

[93] T. Insel *et al.*, "Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders," *Am. J. Psychiatry Online*, no. July, pp. 748–751, 2010.

[94] T. A. Widiger and D. B. Samuel, "Diagnostic categories or dimensions? A question for the Diagnostic and Statistical Manual of Mental Disorders - Fifth Edition," *J. Abnorm. Psychol.*, vol. 114, no. 4, pp. 494–504, 2005.

[95] S. Koh, G. M. Cattell, D. M. Cochran, A. Krasner, F. J. P. Langheim, and D. A. Sasso, "Psychiatrists' use of electronic communication and social media and a proposed framework for future guidelines," *J. Psychiatr. Pract.*, vol. 19, no. 3, pp. 254–263, 2013.

[96] L. W. Roberts, S. Chan, and J. Torous, "New tests, new tools: mobile and connected technologies in advancing psychiatric diagnosis," *npj Digit. Med.*, vol. 1, no. 1, pp. 1–4,

Dec. 2018.

[97] W. van Ballegooijen, H. Riper, P. Cuijpers, P. van Oppen, and J. H. Smit, “Validation of online psychometric instruments for common mental health disorders: A systematic review,” *BMC Psychiatry*, vol. 16, no. 1, 2016.

[98] J. Hunter, K. Corcoran, S. Leeder, and K. Phelps, “Is it time to abandon paper? the use of emails and the Internet for health services research - A cost-effectiveness and qualitative study,” *J. Eval. Clin. Pract.*, vol. 19, no. 5, pp. 855–861, Jun. 2013.

[99] S. Alfonsson, P. Maathz, and T. Hursti, “Interformat reliability of digital psychiatric self-report questionnaires: A systematic review,” *J. Med. Internet Res.*, vol. 16, no. 12, p. e268, Dec. 2014.

[100] I. M. Marks, M. Kenwright, M. McDonough, M. Whittaker, and D. Mataix-Cols, “Saving clinicians’ time by delegating routine aspects of therapy to a computer: A randomized controlled trial in phobia/panic disorder,” *Psychol. Med.*, vol. 34, no. 1, pp. 9–17, Jan. 2004.

[101] P. Farvolden, C. McBride, R. M. Bagby, and P. Ravitz, “A Web-based screening instrument for depression and anxiety disorders in primary care,” *J. Med. Internet Res.*, vol. 5, no. 3, pp. 96–109, 2003.

[102] K. D. Rogers, A. Young, K. Lovell, M. Campbell, P. R. Scott, and S. Kendal, “The British sign language versions of the patient health questionnaire, the generalized anxiety disorder 7-item scale, and the work and social adjustment scale,” *J. Deaf Stud. Deaf Educ.*, vol. 18, no. 1, pp. 110–122, 2013.

[103] W. van Ballegooijen, H. Riper, T. Donker, K. Martin Abello, I. Marks, and P. Cuijpers, “Single-item screening for agoraphobic symptoms: Validation of a web-based audiovisual screening instrument,” *PLoS One*, vol. 7, no. 7, p. e38480, Jul. 2012.

[104] N. Fekadu, W. Shibeshi, and E. Engidawork, “Major Depressive Disorder: Pathophysiology and Clinical Management,” *J. Depress. Anxiety*, vol. 06, no. 01, p. 255, 2017.

[105] M. Shadrina, E. A. Bondarenko, and P. A. Slominsky, “Genetics factors in major depression disease,” *Front. Psychiatry*, vol. 9, no. JUL, p. 334, 2018.

[106] E. Jesulola, P. Micalos, and I. J. Baguley, “Understanding the pathophysiology of depression: From monoamines to the neurogenesis hypothesis model - are we there yet?,” *Behav. Brain Res.*, vol. 341, pp. 79–90, Apr. 2018.

[107] D. H. Geschwind and J. Flint, “Genetics and genomics of psychiatric disease,” *Science (80-.).*, vol. 349, no. 6255, pp. 177–180, 2010.

[108] P. F. Sullivan, F. C. Michael Neale, and K. S. Kendler, “Genetic Epidemiology of Major Depression: Review and Meta-Analysis,” *Am. J. Psychiatry*, vol. 157, p. 10, 2000.

[109] K. S. Kendler, C. O. Gardner, M. C. Neale, and C. A. Prescott, “Genetic risk factors for major depression in men and women: Similar or different heritabilities and same or partly distinct genes?,” *Psychol. Med.*, vol. 31, no. 4, pp. 605–616, 2001.

[110] Daniel, K. S. Nishijima David L, W. David H, and H. James F, “Genetic studies of major depressive disorder: Why are there no GWAS findings, and what can we do about it?,” *Biol. Psychiatry*, vol. 76, no. 7, pp. 510–512, 2014.

[111] P. F. Sullivan *et al.*, “A mega-Analysis of genome-wide association studies for major depressive disorder,” *Mol. Psychiatry*, vol. 18, no. 4, pp. 497–511, Apr. 2013.

[112] K. Hek *et al.*, “A genome-wide association study of depressive symptoms,” *Biol. Psychiatry*, vol. 73, no. 7, pp. 667–678, 2013.

[113] N. R. Wray *et al.*, “Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression,” *Nat. Genet.*, vol. 50, no. 5, pp. 668–681, May 2018.

[114] D. M. Howard *et al.*, “Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions,” *Nat. Neurosci.*, vol. 22, no. 3, pp. 343–352, Mar. 2019.

[115] D. P. Chapman, C. L. Whitfield, V. J. Felitti, S. R. Dube, V. J. Edwards, and R. F. Anda, “Adverse childhood experiences and the risk of depressive disorders in adulthood,” *J. Affect. Disord.*, vol. 82, no. 2, pp. 217–225, 2004.

[116] V. J. Felitti *et al.*, “Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults: The Adverse Childhood Experiences (ACE) Study,” *Am. J. Prev. Med.*, vol. 14, no. 4, pp. 245–258, 1998.

[117] E. C. Nelson *et al.*, “Association between self-reported childhood sexual abuse and adverse psychosocial outcomes: Results from a twin study,” *Arch. Gen. Psychiatry*, vol. 59, no. 2, pp. 139–145, Feb. 2002.

[118] R. C. Kessler, “The Effect of Stressful Life Events On Depression,” *Annu. Rev. Psychol.*, vol. 48, no. 1, pp. 191–214, Feb. 1997.

[119] V. Lorant, D. Deliège, W. Eaton, A. Robert, P. Philippot, and M. Ansseau, “Socioeconomic inequalities in depression: A meta-analysis,” *Am. J. Epidemiol.*, vol. 157, no. 2, pp. 98–112, 2003.

[120] M. T. Tsuang, T. Francis, K. Minor, A. Thomas, and W. S. Stone, “Genetics of smoking and depression,” *Hum. Genet.*, vol. 131, no. 6, pp. 905–915, Jun. 2012.

[121] M. Lyons *et al.*, “A twin study of smoking, nicotine dependence, and major depression in men,” *Nicotine Tob Res*, vol. 10, no. 1, pp. 97–108, Jan. 2008.

[122] M. A. Crocq, “Alcohol, nicotine, caffeine, and mental disorders,” *Dialogues Clin. Neurosci.*, vol. 5, no. 2, pp. 175–185, 2003.

[123] J. C. Haynes *et al.*, “Alcohol consumption as a risk factor for anxiety and depression,” *Br. J. Psychiatry*, vol. 187, no. 6, pp. 544–551, 2005.

[124] D. M. Fergusson, J. M. Boden, and L. J. Horwood, “Tests of causal links between alcohol abuse or dependence and major depression,” *Arch. Gen. Psychiatry*, vol. 66, no. 3, pp. 260–266, 2009.

[125] A. Saban and A. J. Flisher, “The Association between Psychopathology and Substance Use in Young People: A Review of the Literature,” *J. Psychoactive Drugs*, vol. 42, no. 1, pp. 37–47, 2010.

[126] J. Copeland, S. Rooke, and W. Swift, “Changes in cannabis use among young people,” *Curr. Opin. Psychiatry*, vol. 26, no. 4, pp. 325–329, Jul. 2013.

[127] W. L. Mangerud, O. Bjerkeset, T. L. Holmen, S. Lydersen, and M. S. Indredavik, “Smoking, alcohol consumption, and drug use among adolescents with psychiatric disorders compared with a population based sample,” *J. Adolesc.*, vol. 37, no. 7, pp. 1189–1199, 2014.

[128] F. B. Schuch *et al.*, “Physical activity and incident depression: A meta-analysis of prospective cohort studies,” *Am. J. Psychiatry*, vol. 175, no. 7, pp. 631–648, 2018.

[129] F. Schuch *et al.*, “Physical activity and sedentary behavior in people with major depressive disorder: A systematic review and meta-analysis,” *J. Affect. Disord.*, vol. 210, pp. 139–150, Mar. 2017.

[130] C. Hryhorczuk, S. Sharma, and S. E. Fulton, “Metabolic disturbances connecting obesity and depression,” *Front. Neurosci.*, vol. 7, no. 7 OCT, p. 177, Oct. 2013.

[131] F. S. Luppino *et al.*, “Overweight, obesity, and depression: A systematic review and meta-analysis of longitudinal studies,” *Arch. Gen. Psychiatry*, vol. 67, no. 3, pp. 220–229, Mar. 2010.

[132] R. Rosmond, “Obesity and depression: Same disease, different names?,” *Med. Hypotheses*, vol. 62, no. 6, pp. 976–979, 2004.

[133] B. Bondy, “Pathophysiology of depression and mechanisms of treatment,” *Dialogues Clin. Neurosci.*, vol. 4, no. 1, pp. 7–20, Mar. 2002.

[134] P. L. Delgado, “Depression: The case for a monoamine deficiency,” *J. Clin. Psychiatry*, vol. 61, no. SUPPL. 6, pp. 7–11, 2000.

[135] V. Krishnan and E. J. Nestler, “The molecular neurobiology of depression,” *Nature*, vol. 455, no. 7215, pp. 894–902, 2008.

[136] O. Berton and E. J. Nestler, “New approaches to antidepressant drug discovery: Beyond monoamines,” *Nature Reviews Neuroscience*, vol. 7, no. 2. pp. 137–151, Feb-2006.

[137] C. Pittenger and R. S. Duman, “Stress, depression, and neuroplasticity: A convergence of mechanisms,” *Neuropsychopharmacology*, vol. 33, no. 1, pp. 88–109, Jan. 2008.

[138] R. H. Belmaker, “Bipolar disorder,” *N. Engl. J. Med.*, vol. 351, no. 5, pp. 476-486+518, Jul. 2004.

[139] M. Julio-Pieper and T. G. Dinan, “The hypothalamic-pituitary-adrenal axis in depression,” *Depress. From Psychopathol. to Pharmacother.*, vol. 27, pp. 20–31, 2010.

[140] C. Stetler and G. E. Miller, “Depression and hypothalamic-pituitary-adrenal activation: A quantitative summary of four decades of research,” *Psychosom. Med.*, vol. 73, no. 2, pp. 114–126, Feb. 2011.

[141] T. W. W. Pace and A. H. Miller, “Cytokines and glucocorticoid receptor signaling: Relevance to major depression,” *Ann. N. Y. Acad. Sci.*, vol. 1179, pp. 86–105, 2009.

[142] F. P. Varghese and E. S. Brown, “The Hypothalamic-Pituitary-Adrenal Axis in Major Depressive Disorder: A Brief Primer for Primary Care Physicians,” *Prim. Care Companion J Clin Psychiatry*, vol. 3, no. 4, 2001.

[143] S. A. Vreeburg *et al.*, “Major Depressive Disorder and Hypothalamic-Pituitary-Adrenal Axis Activity,” *Arch. Gen. Psychiatry*, vol. 66, no. 6, p. 617, 2009.

[144] J. D. Bremner, M. Narayan, E. R. Anderson, L. H. Staib, H. L. Miller, and D. S. Charney, “Hippocampal volume reduction in major depression,” *Am. J. Psychiatry*, vol. 157, no. 1, pp. 115–117, 2000.

[145] M. C. McKinnon, K. Yucel, A. Nazarov, and G. M. MacQueen, “A meta-analysis examining clinical predictors of hippocampal volume in patients with major depressive disorder,” *J. Psychiatry Neurosci.*, vol. 34, no. 1, pp. 41–54, 2009.

[146] C. Heim, D. J. Newport, T. Mletzko, A. H. Miller, and C. B. Nemeroff, “The link between childhood trauma and depression: Insights from HPA axis studies in humans,” *Psychoneuroendocrinology*, vol. 33, no. 6, pp. 693–710, 2008.

[147] A. Cattaneo *et al.*, “Inflammation and neuronal plasticity: a link between childhood trauma and depression pathogenesis,” *Front. Cell. Neurosci.*, vol. 9, p. 40, Mar. 2015.

[148] M. F. Juruena, “Early-life stress and HPA axis trigger recurrent adulthood depression,” *Epilepsy Behav.*, vol. 38, pp. 148–159, Sep. 2014.

[149] E. A. Young, “Sex differences and the HPA axis: implications for psychiatric disease,”

J. Gend. Specif. Med., vol. 1, no. 1, pp. 21–27, Sep. 1998.

[150] C. L. Raison, L. Capuron, and A. H. Miller, "Cytokines sing the blues: inflammation and the pathogenesis of depression," *Trends Immunol.*, vol. 27, no. 1, pp. 24–31, 2006.

[151] M. Maes, "Major depression and activation of the inflammatory response system," *Adv. Exp. Med. Biol.*, vol. 461, pp. 25–46, 1999.

[152] M. Maes *et al.*, "Increased serum IL-6 and IL-1 receptor antagonist concentrations in major depression and treatment resistant depression," *Cytokine*, vol. 9, no. 11, pp. 853–858, 1997.

[153] S. Alesci *et al.*, "Major depression is associated with significant diurnal elevations in plasma interleukin-6 levels, a shift of its circadian rhythm, and loss of physiological complexity in its secretion: Clinical implications," *J. Clin. Endocrinol. Metab.*, vol. 90, no. 5, pp. 2522–2530, 2005.

[154] J. Levine, Y. Barak, K. N. R. Chengappa, A. Rapoport, M. Rebey, and V. Barak, "Cerebrospinal cytokine levels in patients with acute depression," *Neuropsychobiology*, vol. 40, no. 4, pp. 171–176, 1999.

[155] Y. Dowlati *et al.*, "A Meta-Analysis of Cytokines in Major Depression," *Biol. Psychiatry*, vol. 67, no. 5, pp. 446–457, Mar. 2010.

[156] F. M. Schmidt *et al.*, "Cytokine levels in depressed and non-depressed subjects, and masking effects of obesity," *J. Psychiatr. Res.*, vol. 55, no. 1, pp. 29–34, 2014.

[157] A. J. Thomas, S. Davis, C. Morris, E. Jackson, R. Harrison, and J. T. O'Brien, "Increase in interleukin-1β in late-life depression," *Am. J. Psychiatry*, vol. 162, no. 1, pp. 175–177, Jan. 2005.

[158] N. Vogelzangs *et al.*, "Association of depressive disorders, depression characteristics and antidepressant medication with inflammation," *Transl. Psychiatry*, vol. 2, p. 79, 2012.

[159] T. W. W. Pace, F. Hu, and A. H. Miller, "Cytokine-effects on glucocorticoid receptor function: Relevance to glucocorticoid resistance and the pathophysiology and treatment of major depression," *Brain. Behav. Immun.*, vol. 21, no. 1, pp. 9–19, Jan. 2007.

[160] P. A. Zunszain, C. Anacker, A. Cattaneo, and L. A. Carvalho, "Glucocorticoids, cytokines and brain abnormalities in depression," *Prog. Neuro-Psychopharmacology Biol. Psychiatry*, vol. 35, no. 3, pp. 722–729, 2011.

[161] T. W. W. Pace *et al.*, "Increased stress-induced inflammatory responses in male patients with major depression and increased early life stress," *Am. J. Psychiatry*, vol. 163, no. 9, pp. 1630–1633, 2006.

[162] A. Danese, C. M. Pariante, A. Caspi, A. Taylor, and R. Poulton, "Childhood maltreatment predicts adult inflammation in a life-course study," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 4, pp. 1319–1324, 2007.

[163] D. Brites and A. Fernandes, "Neuroinflammation and depression: Microglia activation, extracellular microvesicles and microRNA dysregulation," *Front. Cell. Neurosci.*, vol. 9, no. DEC, pp. 1–20, Dec. 2015.

[164] S. G. Torres-Platas, C. Cruceanu, G. G. Chen, G. Turecki, and N. Mechawar, "Evidence for increased microglial priming and macrophage recruitment in the dorsal anterior cingulate white matter of depressed suicides," *Brain. Behav. Immun.*, vol. 42, pp. 50–59, Nov. 2014.

[165] G. Z. Réus *et al.*, "The role of inflammation and microglial activation in the pathophysiology of psychiatric disorders," *Neuroscience*, vol. 300, pp. 141–154, Aug. 2015.

[166] M. A. Calcia, D. R. Bonsall, P. S. Bloomfield, S. Selvaraj, T. Barichello, and O. D. Howes, "Stress and neuroinflammation: A systematic review of the effects of stress on microglia and the implications for mental illness," *Psychopharmacology (Berl).*, vol. 233, no. 9, pp. 1637–1650, May 2016.

[167] R. Ottman, "Gene-environment interaction: Definitions and study designs," *Prev. Med. (Baltim).*, vol. 25, no. 6, pp. 764–770, Nov. 1996.

[168] N. Lopizzo *et al.*, "Gene-environment interaction in major depression: Focus on experience-dependent biological systems," *Front. Psychiatry*, vol. 6, no. MAY, 2015.

[169] R. Uher, "Gene-environment interactions in common mental disorders: An update and strategy for a genome-wide search," *Soc. Psychiatry Psychiatr. Epidemiol.*, vol. 49, no. 1, pp. 3–14, 2014.

[170] T. Klengel and E. B. Binder, "Epigenetics of Stress-Related Psychiatric Disorders and Gene × Environment Interactions," *Neuron*, vol. 86, no. 6, pp. 1343–1357, 2015.

[171] T. Klengel *et al.*, "Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions," *Nat. Neurosci.*, vol. 16, no. 1, pp. 33–41, Jan. 2013.

[172] P. O. McGowan *et al.*, "Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse," *Nat. Neurosci.*, vol. 12, no. 3, pp. 342–348, 2009.

[173] G. Hasler, "Pathophysiology of depression: Do we have any solid evidence of interest to clinicians?," *World Psychiatry*, vol. 9, no. 3, pp. 155–161, 2010.

[174] J. Maniam, C. Antoniadis, and M. J. Morris, "Early-life stress, HPA axis adaptation, and mechanisms contributing to later health outcomes," *Front. Endocrinol. (Lausanne).*, vol.

5, no. MAY, 2014.

[175] R. S. Duman and L. M. Monteggia, “A Neurotrophic Model for Stress-Related Mood Disorders,” *Biol. Psychiatry*, vol. 59, no. 12, pp. 1116–1127, Jun. 2006.

[176] L. M. Monteggia *et al.*, “Essential role of brain-derived neurotrophic factor in adult hippocampal function,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 29, pp. 10827–10832, 2004.

[177] B. H. Lee and Y. K. Kim, “The roles of BDNF in the pathophysiology of major depression and in antidepressant treatment,” *Psychiatry Investig.*, vol. 7, no. 4, pp. 231–235, 2010.

[178] M. Polyakova, K. Stuke, K. Schuemberg, K. Mueller, P. Schoenknecht, and M. L. Schroeter, “BDNF as a biomarker for successful treatment of mood disorders: A systematic & quantitative meta-analysis,” *J. Affect. Disord.*, vol. 174, pp. 432–440, 2015.

[179] F. Karege, G. Perret, G. Bondolfi, M. Schwald, G. Bertschy, and J. M. Aubry, “Decreased serum brain-derived neurotrophic factor levels in major depressed patients,” *Psychiatry Res.*, vol. 109, no. 2, pp. 143–148, 2002.

[180] F. Karege, G. Bondolfi, N. Gervasoni, M. Schwald, J. M. Aubry, and G. Bertschy, “Low Brain-Derived Neurotrophic Factor (BDNF) levels in serum of depressed patients probably results from lowered platelet BDNF release unrelated to platelet reactivity,” *Biol. Psychiatry*, vol. 57, no. 9, pp. 1068–1072, May 2005.

[181] E. Shimizu *et al.*, “Alterations of serum levels of brain-derived neurotrophic factor (BDNF) in depressed patients with or without antidepressants,” *Biol. Psychiatry*, vol. 54, no. 1, pp. 70–75, Jul. 2003.

[182] B. H. Lee, H. Kim, S. H. Park, and Y. K. Kim, “Decreased plasma BDNF level in depressive patients,” *J. Affect. Disord.*, vol. 101, no. 1–3, pp. 239–244, Aug. 2007.

[183] S. U. Maier, A. B. Makwana, and T. A. Hare, “Acute Stress Impairs Self-Control in Goal-Directed Choice by Altering Multiple Functional Connections within the Brain’s Decision Circuits,” *Neuron*, vol. 87, no. 3, pp. 621–631, 2015.

[184] R. Nusslock and G. E. Miller, “Early-life adversity and physical and emotional health across the lifespan: A neuroimmune network hypothesis,” *Biological Psychiatry*, vol. 80, no. 1. pp. 23–32, 2016.

[185] J. P. Hamilton, D. Ph, M. G. Lemus, R. F. Johnson, I. H. Gotlib, and D. Ph, “Functional Neuroimaging of Major depressive disorder: A Meta-Analysis and New Integration of Baseline Activation and Neural Response data,” *Am. J. Psychiatry*, vol. 169, no. 7, pp. 693–703, 2012.

[186] R. H. Kaiser, J. R. Andrews-Hanna, T. D. Wager, and D. A. Pizzagalli, "Large-scale network dysfunction in major depressive disorder: A meta-analysis of resting-state functional connectivity," *JAMA Psychiatry*, vol. 72, no. 6, pp. 603–611, 2015.

[187] J. P. Hamilton, M. C. Chen, and I. H. Gotlib, "Neural systems approaches to understanding major depressive disorder: An intrinsic functional organization perspective," *Neurobiol. Dis.*, vol. 52, pp. 4–11, Apr. 2013.

[188] S. E. Lakhan, K. Vieira, and E. Hamlat, "Biomarkers in psychiatry: Drawbacks and potential for misuse," *Int. Arch. Med.*, vol. 3, no. 1, 2010.

[189] A. J. Atkinson *et al.*, "Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework," *Clin. Pharmacol. Ther.*, vol. 69, no. 3, pp. 89–95, Mar. 2001.

[190] A. Halfin, "Depression: The benefits of early and appropriate treatment," *Am. J. Manag. Care*, vol. 13, no. 11 SUPPL. 4, pp. S92–S97, 2007.

[191] N. Carlomagno *et al.*, "Diagnostic, Predictive, Prognostic, and Therapeutic Molecular Biomarkers in Third Millennium: A Breakthrough in Gastric Cancer," *Biomed Res. Int.*, vol. 2017, 2017.

[192] R. H. Perlis, "Betting on biomarkers," *Am. J. Psychiatry*, vol. 168, no. 3, pp. 234–236, Mar. 2011.

[193] A. L. Comes, S. Papiol, T. Mueller, P. E. Geyer, M. Mann, and T. G. Schulze, "Proteomics for blood biomarker exploration of severe mental illness: pitfalls of the past and potential for the future," *Transl. Psychiatry*, vol. 8, 2018.

[194] A. G. Ngounou Wetie *et al.*, "Mass spectrometry for the detection of potential psychiatric biomarkers," *J. Mol. psychiatry*, vol. 1, no. 1, p. 8, 2013.

[195] S. Patel, "Role of proteomics in biomarker discovery and psychiatric disorders: current status, potentials, limitations and future challenges," *Expert Rev. Proteomics*, vol. 9, no. 3, pp. 249–265, Jun. 2012.

[196] D. Martins-de-Souza, L. W. Harris, P. C. Guest, C. W. Turck, and S. Bahn, "The role of proteomics in depression research," *Eur. Arch. Psychiatry Clin. Neurosci.*, vol. 260, no. 6, pp. 499–506, Sep. 2010.

[197] G. Duan and D. Walther, "The Roles of Post-translational Modifications in the Context of Protein Interaction Networks," *PLoS Comput. Biol.*, vol. 11, no. 2, 2015.

[198] A. H. Smits and M. Vermeulen, "Characterizing Protein–Protein Interactions Using Mass Spectrometry: Challenges and Opportunities," *Trends Biotechnol.*, vol. 34, no. 10, pp. 825–834, 2016.

[199] B. Krastins *et al.*, "Rapid development of sensitive, high-throughput, quantitative and

highly selective mass spectrometric targeted immunoassays for clinically important proteins in human plasma and serum," *Clin. Biochem.*, 2013.

[200] B. B. Haab, "Applications of antibody array platforms," *Curr. Opin. Biotechnol.*, vol. 17, no. 4, pp. 415–421, Aug. 2006.

[201] A. N. Hoofnagle and M. H. Wener, "The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry," 2009.

[202] S. Ozcan *et al.*, "Towards reproducible MRM based biomarker discovery using dried blood spots," *Sci. Rep.*, vol. 7, p. 45178, Mar. 2017.

[203] A. N. Hoofnagle and M. H. Wener, "The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry," *J. Immunol. Methods*, vol. 347, no. 1–2, pp. 3–11, Aug. 2009.

[204] A. Efeyan *et al.*, "Cancer-associated IDH1 mutations produce 2-hydroxyglutarate," *Nature*, 2009.

[205] Z. Maunsell, D. J. Wright, and S. J. Rainbow, "Routine isotope-dilution liquid chromatography-tandem mass spectrometry assay for simultaneous measurement of the 25-hydroxy metabolites of vitamins D2 and D3," *Clin. Chem.*, 2005.

[206] C. A. Crutchfield, S. N. Thomas, L. J. Sokoll, and D. W. Chan, "Advances in mass spectrometry-based clinical biomarker discovery," *Clin. Proteomics*, 2016.

[207] G. L. Glish and R. W. Vachet, "The basics of mass spectrometry in the twenty-first century," *Nat. Rev. Drug Discov.*, vol. 2, no. 2, pp. 140–150, Feb. 2003.

[208] E. Boschetti, M. C. M. Chung, and P. G. Righetti, "'The quest for biomarkers': Are we on the right technical track?," *Proteomics - Clinical Applications*. The Economist, 2012.

[209] C. E. Parker and C. H. Borchers, "Mass spectrometry based biomarker discovery, verification, and validation - Quality assurance and control of protein biomarker assays," *Mol. Oncol.*, vol. 8, no. 4, pp. 840–858, Jun. 2014.

[210] D. C. Liebler and L. J. Zimmerman, "Targeted quantitation of proteins by mass spectrometry," *Biochemistry*, vol. 52, no. 22, pp. 3797–806, Jun. 2013.

[211] W. J. Griffiths and Y. Wang, "Mass spectrometry: From proteomics to metabolomics and lipidomics," *Chem. Soc. Rev.*, vol. 38, no. 7, pp. 1882–1896, 2009.

[212] E. P. Diamandis, "Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: Opportunities and potential limitations," *Mol. Cell. Proteomics*, vol. 3, no. 4, pp. 367–378, Apr. 2004.

[213] S. Patel, "Role of proteomics in biomarker discovery: Prognosis and diagnosis of neuropsychiatric disorders," in *Advances in Protein Chemistry and Structural Biology*,

vol. 94, Academic Press Inc., 2014, pp. 39–75.

[214] T. H. Tsai *et al.*, "LC-MS/MS-based serum proteomics for identification of candidate biomarkers for hepatocellular carcinoma," *Proteomics*, vol. 15, no. 13, pp. 2369–2381, 2015.

[215] P. Picotti and R. Aebersold, "Selected reaction monitoring-based proteomics: Workflows, potential, pitfalls and future directions," *Nat. Methods*, vol. 9, no. 6, pp. 555–566, 2012.

[216] Y. Hacimusalar and E. Eşel, "Suggested biomarkers for major depressive disorder," *Arch. Neuropsychiatry*, vol. 55, no. 3, pp. 280–290, 2018.

[217] N. L. G. L. Anderson and N. L. G. L. Anderson, "The Human Plasma Proteome: History, Character, and Diagnostic Prospects," *Mol. Cell. Proteomics*, vol. 1, no. 11, pp. 845–867, 2002.

[218] A. Alawieh *et al.*, "Systems Biology, Bioinformatics, and Biomarkers in Neuropsychiatry," *Front. Neurosci.*, vol. 6, p. 187, 2012.

[219] J. N. Adkins *et al.*, "Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry," *Mol. Cell. Proteomics*, vol. 1, no. 12, pp. 947–955, 2002.

[220] M. L. Wong *et al.*, "Pronounced and sustained central hypernoradrenergic function in major depression with melancholic features: relation to hypercortisolism and corticotropin-releasing hormone," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 1, pp. 325–30, Jan. 2000.

[221] F. Kaestner *et al.*, "Different activation patterns of proinflammatory cytokines in melancholic and non-melancholic major depression are associated with HPA axis activity," *J. Affect. Disord.*, vol. 87, no. 2–3, pp. 305–311, Aug. 2005.

[222] M. K. Chan, J. D. Cooper, M. Bot, J. Steiner, B. W. J. H. Penninx, and S. Bahn, "Identification of an Immune-Neuroendocrine Biomarker Panel for Detection of Depression: A Joint Effects Statistical Approach," *Neuroendocrinology*, vol. 103, no. 6, pp. 693–710, 2016.

[223] G. I. Papakostas *et al.*, "Assessment of a multi-assay, serum-based biological diagnostic test for major depressive disorder: a pilot and replication study," *Mol. Psychiatry*, vol. 18, no. 3, pp. 332–9, Mar. 2013.

[224] T. T. G. Dinan and T. T. G. Dinan, "Inflammatory markers in depression," *Curr. Opin. Psychiatry*, vol. 22, no. 1, pp. 32–36, Jan. 2009.

[225] R. Strawbridge, A. H. Young, and A. J. Cleare, "Biomarkers for depression: Recent

insights, current challenges and future prospects," *Neuropsychiatr. Dis. Treat.*, vol. 13, pp. 1245–1262, 2017.

[226] V. Stelzhammer *et al.*, "Proteomic changes in serum of first onset, antidepressant drug-naïve major depression patients," *Int. J. Neuropsychopharmacol.*, vol. 17, no. 10, pp. 1–10, Oct. 2014.

[227] E. Domenici *et al.*, "Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections," *PLoS One*, vol. 5, no. 2, p. e9166, Feb. 2010.

[228] R. L. Preece, S. Y. S. Han, and S. Bahn, "Proteomic approaches to identify blood-based biomarkers for depression and bipolar disorders," *Expert Rev. Proteomics*, vol. 15, no. 4, pp. 325–340, 2018.

[229] A. Parekh, D. Smeeth, Y. Milner, S. Thuret, and S. Thure, "The Role of Lipid Biomarkers in Major Depression," *Healthcare*, vol. 5, no. 1, p. 5, Feb. 2017.

[230] S. M. Hanash, S. M. Hanash, C. S. Baik, and O. Kallioniemi, "Emerging molecular biomarkers - blood-based strategies to detect and monitor cancer," *Nat. Publ. Gr.*, vol. 8, pp. 142–150, 2011.

[231] V. Kulasingam and E. P. Diamandis, "Strategies for discovering novel cancer biomarkers through utilization of emerging technologies," *Nat. Clin. Pract. Oncol.*, vol. 5, no. 10, pp. 588–599, Aug. 2008.

[232] M. Dowsett and A. K. Dunbier, "Emerging biomarkers and new understanding of traditional markers in personalized therapy for breast cancer," *Clin. Cancer Res.*, vol. 14, no. 24, pp. 8019–8026, Dec. 2008.

[233] M. P. Pavlou, E. P. Diamandis, and I. M. Blasutig, "The Long Journey of Cancer Biomarkers from the Bench to the Clinic," *Clin. Chem.*, vol. 59, no. 1, pp. 147–157, Jan. 2013.

[234] J. W. Day and W. M. Rosenberg, "The enhanced liver fibrosis (ELF) test in diagnosis and management of liver fibrosis," *Br. J. Hosp. Med.*, vol. 79, no. 12, pp. 694–699, Dec. 2018.

[235] F. Kobeissy, A. Alawieh, S. Mondello, R. M. Boustany, and M. S. Gold, "Biomarkers in psychiatry: How close are we?," *Front. Psychiatry*, vol. 3, no. JAN, p. 187, 2013.

[236] B. Hernández, S. R. Pennington, and A. C. Parnell, "Bayesian methods for proteomic biomarker development," *EuPA Open Proteomics*, vol. 9, pp. 54–64, 2015.

[237] H. D. Schmidt, R. C. Shelton, and R. S. Duman, "Functional biomarkers of depression: Diagnosis, treatment, and pathophysiology," *Neuropsychopharmacology*, vol. 36, no.

12, pp. 2375–2394, 2011.

[238] H.-B. B. Xu *et al.*, “Comparative proteomic analysis of plasma from major depressive patients: identification of proteins associated with lipid metabolism and immunoregulation,” *Int. J. Neuropsychopharmacol.*, vol. 15, no. 10, pp. 1413–1425, Nov. 2012.

[239] M. Bot *et al.*, “Serum proteomic profiling of major depressive disorder,” *Transl. Psychiatry*, vol. 5, no. 7, p. e599, 2015.

[240] J. Lee *et al.*, “Proteomic Analysis of Serum from Patients with Major Depressive Disorder to Compare Their Depressive and Remission Statuses,” *Psychiatry Investig.*, vol. 12, no. 2, p. 249, Apr. 2015.

[241] M. Y. Lee *et al.*, “Discovery of serum protein biomarkers in drug-free patients with major depressive disorder,” *Prog. Neuropsychopharmacol. Biol. Psychiatry*, vol. 69, pp. 60–68, 2016.

[242] Q. Wang *et al.*, “ITRAQ technology-based identification of human peripheral serum proteins associated with depression,” *Neuroscience*, vol. 330, pp. 291–325, 2016.

[243] E. Y. Kim, M. Y. Lee, S. H. Kim, K. Ha, K. P. Kim, and Y. M. Ahn, “Diagnosis of major depressive disorder by combining multimodal information from heart rate dynamics and serum proteomics using machine-learning algorithm,” *Prog. Neuro-Psychopharmacology Biol. Psychiatry*, vol. 76, pp. 65–71, Jun. 2017.

[244] G. Kang *et al.*, “Fibrinogen and kininogen are potential serum protein biomarkers for depressive disorder,” *Clin. Lab.*, vol. 65, no. 10, pp. 1887–1893, 2019.

[245] J. A. Bilello *et al.*, “MDDScore: Confirmation of a blood test to aid in the diagnosis of major depressive disorder,” *J. Clin. Psychiatry*, vol. 76, no. 2, pp. e199–e206, Feb. 2015.

[246] S. H. Lee *et al.*, “Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs,” *Nat. Genet.*, vol. 45, no. 9, pp. 984–994, Sep. 2013.

[247] J. S. W. Smoller *et al.*, “Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis,” *Lancet*, vol. 381, no. 9875, pp. 1371–1379, Apr. 2013.

[248] T. G. Schulze *et al.*, “Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder,” *World J. Biol. Psychiatry*, vol. 15, no. 3, pp. 200–208, 2014.

[249] N. Yuan, Y. Chen, Y. Xia, J. Dai, and C. Liu, “Inflammation-related biomarkers in major psychiatric disorders: a cross-disorder assessment of reproducibility and specificity in 43 meta-analyses,” *Transl. Psychiatry*, vol. 9, no. 1, pp. 1–13, Dec. 2019.

[250] N. Müller, “Inflammation in schizophrenia: Pathogenetic aspects and therapeutic

considerations," *Schizophr. Bull.*, vol. 44, no. 5, pp. 973–982, Aug. 2018.

[251] A. Muneer, "Bipolar disorder: Role of inflammation and the development of disease biomarkers," *Psychiatry Investig.*, vol. 13, no. 1, pp. 18–33, Jan. 2016.

[252] H. Zhao and D. R. Nyholt, "Gene-based analyses reveal novel genetic overlap and allelic heterogeneity across five major psychiatric disorders," *Hum. Genet.*, vol. 136, no. 2, pp. 263–274, 2017.

[253] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement," *BMJ*, vol. 350, Jan. 2015.

[254] J. M. T. Hendriksen, G. J. Geersing, K. G. M. Moons, and J. A. H. de Groot, "Diagnostic and prognostic prediction models," *J. Thromb. Haemost.*, vol. 11, pp. 129–141, Jun. 2013.

[255] M. Pavlou *et al.*, "How to develop a more accurate risk prediction model when there are few events," *BMJ*, vol. 351, 2015.

[256] M. E. Shipe, S. A. Deppen, F. Farjah, and E. L. Grogan, "Developing prediction models for clinical use using logistic regression: An overview," *J. Thorac. Dis.*, vol. 11, pp. S574–S584, 2019.

[257] K. G. M. M. Moons, P. Royston, Y. Vergouwe, D. E. Grobbee, and D. G. Altman, "Prognosis and prognostic research: What, why, and how?," in *BMJ (Online)*, 23-Feb-2009, vol. 338, no. 7706, pp. 1317–1320.

[258] K. G. M. Moons *et al.*, "Risk prediction models: II. External validation, model updating, and impact assessment," *Heart*, vol. 98, no. 9, pp. 691–698, May 2012.

[259] A. C. Justice, K. E. Covinsky, and J. A. Berlin, "Assessing the generalizability of prognostic information," *Ann. Intern. Med.*, vol. 130, no. 6, pp. 515–524, 1999.

[260] D. Krstajic, L. J. Buturovic, D. E. Leahy, and S. Thomas, "Cross-validation pitfalls when selecting and assessing regression and classification models," *J. Cheminform.*, vol. 6, no. 1, p. 10, Dec. 2014.

[261] B. Hernández, A. Parnell, S. R. Pennington, B. Hern??ndez, A. Parnell, and S. R. Pennington, "Why have so few proteomic biomarkers 'survived' validation? (Sample size and independent validation considerations)," *Proteomics*, vol. 14, no. 13–14, pp. 1587–1592, Jul. 2014.

[262] H. Mischak *et al.*, "Implementation of proteomic biomarkers: Making it work," *Eur. J. Clin. Invest.*, vol. 42, no. 9, pp. 1027–1036, Sep. 2012.

[263] N. Rifai, M. A. Gillette, and S. A. Carr, "Protein biomarker discovery and validation:

the long and uncertain path to clinical utility," *Nat. Biotechnol.*, vol. 24, no. 8, pp. 971–983, Aug. 2006.

[264] E. P. Diamandis, "The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem?," *BMC Med.*, vol. 10, 2012.

[265] M. Vidal *et al.*, "The human proteome - a scientific opportunity for transforming diagnostics, therapeutics, and healthcare," *Clin. Proteomics*, vol. 9, no. 1, p. 6, Jul. 2012.

[266] M. A. Pourhoseingholi, A. R. Baghestani, and M. Vahedi, "How to control confounding effects by statistical analysis," *Gastroenterol. Hepatol. from Bed to Bench*, vol. 5, no. 2, pp. 79–83, 2012.

[267] R. McNamee, "Regression modelling and other methods to control confounding," *Occup. Environ. Med.*, vol. 62, no. 7, pp. 500–506, 2005.

[268] M. J. Marton and R. Weiner, "Practical guidance for implementing predictive biomarkers into early phase clinical studies," *Biomed Res. Int.*, vol. 2013, p. 891391, Oct. 2013.

[269] E. W. Steyerberg *et al.*, "Assessing the performance of prediction models: A framework for some traditional and novel measures," *Epidemiology*, vol. 21, no. 1, pp. 128–138, Jan. 2010.

[270] W. Bouwmeester *et al.*, "Reporting and Methods in Clinical Prediction Research: A Systematic Review," *PLoS Med.*, vol. 9, no. 5, p. e1001221, May 2012.

[271] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstem, "A simulation study of the number of events per variable in logistic regression analysis," *J. Clin. Epidemiol.*, vol. 49, no. 12, pp. 1373–1379, Dec. 1996.

[272] E. W. Steyerberg, M. J. C Eijkemans, F. E. Harrell Jr, and J. F. Habbema, "Prognostic Modeling with Logistic Regression Analysis: In Search of a Sensible Strategy in Small Data Sets," *Med. Decis. Mak.*, vol. 21, no. 1, pp. 45–56, 2001.

[273] M. Pavlou, G. Ambler, S. Seaman, M. De iorio, and R. Z. Omar, "Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events," *Stat. Med.*, vol. 35, no. 7, pp. 1159–1177, Mar. 2016.

[274] K. G. M. Moons *et al.*, "Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker," *Heart*, vol. 98, no. 9, pp. 683–690, May 2012.

[275] M. van Smeden *et al.*, "Sample size for binary logistic prediction models: Beyond events per variable criteria," *Stat. Methods Med. Res.*, vol. 28, no. 8, pp. 2455–2474, Aug. 2019.

[276] E. W. Steyerberg, M. J. C. Eijkemans, F. E. Harrell, and J. D. F. Habbema, "Prognostic

modelling with logistic regression analysis: A comparison of selection and estimation methods in small data sets," *Stat. Med.*, vol. 19, no. 8, pp. 1059–1079, Apr. 2000.

[277] G. Ambler, A. R. Brady, and P. Royston, "Simplifying a prognostic model: A simulation study based on clinical data," *Stat. Med.*, vol. 21, no. 24, pp. 3803–3822, Dec. 2002.

[278] J. Kaiser, "Biomarker tests need closer scrutiny, IOM concludes," *Science (80-.).*, vol. 335, no. 6076, p. 1554, 2012.

[279] G. Ambler, S. Seaman, and R. Z. Omar, "An evaluation of penalised survival methods for developing prognostic models with rare events," *Stat. Med.*, vol. 31, no. 11–12, pp. 1150–1161, 2012.

[280] S. Le Cessie and J. C. Van Houwelingen, "Ridge Estimators in Logistic Regression," *Appl. Stat.*, vol. 41, no. 1, p. 191, 1992.

[281] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[282] E. W. Steyerberg, F. E. Harrell, G. J. J. . Borsboom, M. J. . Eijkemans, Y. Vergouwe, and J. D. F. Habbema, "Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis," *J. Clin. Epidemiol.*, vol. 54, no. 8, pp. 774–781, Aug. 2001.

[283] K. P. Burnham and D. R. Anderson, "Multimodel Inference: Understanding AIC and BIC in Model Selection," *Sociol. Methods Res.*, vol. 33, no. 2, pp. 261–304, 2004.

[284] K. P. Burnham and D. R. Anderson, *Model Selection and Multi-Model Inference: A Practical Information-Theoretical Approach*, Second. Springer, 2002.

[285] S. Roberts and M. A. Martin, "Bootstrap-after-bootstrap model averaging for reducing model uncertainty in model selection for air pollution mortality studies," *Environ. Health Perspect.*, vol. 118, no. 1, pp. 131–136, Jan. 2010.

[286] M. Clyde, "Model averaging," in *Subjective and Objective Bayesian Statistics*, Second Edi., John Wiley & Sons, Inc., 2002, pp. 636–642.

[287] D. Draper, "Assessment and Propagation of Model Uncertainty," *J. R. Stat. Soc. Ser. B*, vol. 57, no. 1, pp. 45–70, 1995.

[288] M. Clyde, "Model uncertainty and health effect studies for particulate matter," *Environmetrics*, vol. 11, no. 6, pp. 745–763, 2000.

[289] M. R. E. Symonds and A. Moussalli, "A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion," *Behav. Ecol. Sociobiol.*, vol. 65, no. 1, pp. 13–21, 2011.

[290] D. Posada and T. R. Buckley, "Model selection and model averaging in phylogenetics:

Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests," *Syst. Biol.*, vol. 53, no. 5, pp. 793–808, 2004.

[291] K. A. Phillips, S. Van Bebber, and A. M. Issa, "Diagnostics and biomarker development: Priming the pipeline," *Nat. Rev. Drug Discov.*, vol. 5, no. 6, pp. 463–469, Jun. 2006.

[292] A. K. Füzéry, J. Levin, M. M. Chan, and D. W. Chan, "Translation of proteomic biomarkers into FDA approved cancer diagnostics: Issues and challenges," *Clin. Proteomics*, vol. 10, no. 1, p. 13, Dec. 2013.

[293] T. G. Lively, M. C. Sachs, and B. A. Conley, "Translating Promising Biomarkers into Oncology Clinical Practice," *Oncol. Hematol. Rev.*, vol. 11, no. 01, p. 14, 2015.

[294] J. D. Cooper *et al.*, "Multimodel inference for biomarker development: an application to schizophrenia," *Transl. Psychiatry*, vol. 9, no. 1, p. 83, 2019.

[295] S. Y. S. Han, J. D. Cooper, S. Ozcan, N. Rustogi, B. W. J. H. Penninx, and S. Bahn, "Integrating proteomic, sociodemographic and clinical data to predict future depression diagnosis in subthreshold symptomatic individuals," *Transl. Psychiatry*, vol. 9, no. 1, p. 277, 2019.

[296] T. Olmert *et al.*, "A Combined Digital and Biomarker Diagnostic Aid for Mood Disorders: The Delta Trial, an Observational Study," *JMIR Res. Protoc.*, vol. 9, no. 6, p. e18453, Aug. 2020.

[297] S. Y. S. Han *et al.*, "Diagnostic prediction model development using data from dried blood spot proteomics and a digital mental health assessment to identify major depressive disorder among individuals presenting with low mood," *Brain. Behav. Immun.*, vol. 90, no. November, pp. 184–195, 2020.

[298] World Medical Association, "World Medical Association declaration of Helsinki: Ethical principles for medical research involving human subjects," *JAMA*, vol. 310, no. 20, pp. 2191–2194, Nov. 2013.

[299] R. S. Tirumalai, K. C. Chan, D. R. A. Prieto, H. J. Issaq, T. P. Conrads, and T. D. Veenstra, "Characterization of the low molecular weight human serum proteome," *Mol. Cell. Proteomics*, vol. 2, no. 10, pp. 1096–1103, Oct. 2003.

[300] J. M. Jacobs *et al.*, "Utilizing Human Blood Plasma for Proteomic Biomarker Discovery," *J. Proteome Res.*, 2005.

[301] J. A. J. J. Jaros, P. C. Guest, S. Bahn, and D. Martins-de-Souza, "Affinity depletion of plasma and serum for mass spectrometry-based proteome analysis," *Methods Mol. Biol.*, vol. 1002, pp. 1–11, 2013.

[302] L. Urbas *et al.*, "Depletion of high-abundance proteins from human plasma using a

combination of an affinity and pseudo-affinity column," *J. Chromatogr. A*, 2009.

[303] P. A. Demirev, "Dried Blood Spots: Analysis and Applications," *Anal. Chem.*, vol. 85, no. 2, pp. 779–789, Jan. 2013.

[304] T. W. McDade, S. A. (Sharon A. A. Williams, and J. J. Snodgrass, "What a Drop Can Do: Dried Blood Spots as a Minimally Invasive Method for Integrating Biomarkers Into Population-Based Research," *Demography*, vol. 44, no. 4, pp. 899–925, Nov. 2007.

[305] V. Ignjatovic, J. Pitt, P. Monagle, and J. M. Craig, "The utility of dried blood spots for proteomic studies: Looking forward to looking back," *Proteomics - Clin. Appl.*, vol. 8, no. 11–12, pp. 896–900, Dec. 2014.

[306] O. P. Trifonova, D. L. Maslov, E. E. Balashova, and P. G. Lokhov, "Evaluation of dried blood spot sampling for clinical metabolomics: Effects of different papers and sample storage stability," *Metabolites*, vol. 9, no. 11, 2019.

[307] R. Guthrie and A. Susi, "A Simple Phenylalanine Method For Detecting Phenylketonuria In Large Populations Of Newborn Infants," *Pediatrics*, vol. 32, pp. 338–343, 1963.

[308] A. J. Wilhelm, J. C. G. den Burger, and E. L. Swart, "Therapeutic Drug Monitoring by Dried Blood Spot: Progress to Date and Future Directions," *Clin. Pharmacokinet.*, vol. 53, no. 11, pp. 961–973, 2014.

[309] S. S. Solomon *et al.*, "Dried blood spots (DBS): A valuable tool for HIV surveillance in developing/tropical countries," *Int. J. STD AIDS*, vol. 13, no. 1, pp. 25–28, Jan. 2002.

[310] L. C. Martial *et al.*, "Dried Blood Spot sampling in psychiatry: Perspectives for improving therapeutic drug monitoring," *Eur. Neuropsychopharmacol.*, vol. 27, no. 3, pp. 205–216, Mar. 2017.

[311] J. Zweigenbaum and J. Henion, "Bioanalytical high-throughput selected reaction monitoring-LC/MS determination of selected estrogen receptor modulators in human plasma: 2000 samples/day," *Anal. Chem.*, vol. 72, no. 11, pp. 2446–2454, Jun. 2000.

[312] N. L. Anderson, M. Razavi, T. W. Pearson, G. Kruppa, R. Paape, and D. Suckau, "Precision of heavy-light peptide ratios measured by MALDI-TOF mass spectrometry," in *Journal of Proteome Research*, 2012, vol. 11, no. 3, pp. 1868–1878.

[313] L. Anderson and C. L. Hunter, "Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins," *Mol. Cell. Proteomics*, vol. 5, no. 4, pp. 573–588, Apr. 2006.

[314] D. Domanski *et al.*, "MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma," *Proteomics*, vol. 12, no. 8, pp. 1222–1243, Apr.

2012.

[315] A. G. Chambers, A. J. Percy, J. Yang, A. G. Camenzind, and C. H. Borchers, "Multiplexed Quantitation of Endogenous Proteins in Dried Blood Spots by Multiple Reaction Monitoring - Mass Spectrometry," *Mol. Cell. Proteomics*, vol. 12, no. 3, pp. 781–791, Mar. 2013.

[316] A. G. Chambers, A. J. Percy, J. Yang, and C. H. Borchers, "Multiple Reaction Monitoring Enables Precise Quantification of 97 Proteins in Dried Blood Spots," *Mol. Cell. Proteomics*, vol. 14, no. 11, pp. 3094–3104, Nov. 2015.

[317] Y. V. Karpievitch, A. D. Polpitiya, G. A. Anderson, R. D. Smith, and A. R. Dabney, "Liquid chromatography mass spectrometry-based proteomics: Biological and technological aspects," *Ann. Appl. Stat.*, vol. 4, no. 4, pp. 1797–1823, 2010.

[318] H. Rodriguez *et al.*, "Reconstructing the pipeline by introducing multiplexed multiple reaction monitoring mass spectrometry for cancer biomarker verification: An NCI-CPTC initiative perspective," *Proteomics - Clin. Appl.*, vol. 4, no. 12, pp. 904–914, 2010.

[319] E. H. Wang, P. C. Combe, and K. A. Schug, "Multiple Reaction Monitoring for Direct Quantitation of Intact Proteins Using a Triple Quadrupole Mass Spectrometer," *J. Am. Soc. Mass Spectrom.*, vol. 27, no. 5, pp. 886–896, May 2016.

[320] S. E. Ong and M. Mann, "Mass Spectrometry–Based Proteomics Turns Quantitative," *Nat. Chem. Biol.*, vol. 1, no. 5, pp. 252–262, Sep. 2005.

[321] A. N. Hoofnagle *et al.*, "Recommendations for the generation, quantification, storage, and handling of peptides used for mass spectrometry-based assays," *Clin. Chem.*, vol. 62, no. 1, pp. 48–69, 2016.

[322] B. MacLean *et al.*, "Skyline: An open source document editor for creating and analyzing targeted proteomics experiments," *Bioinformatics*, vol. 26, no. 7, pp. 966–968, Apr. 2010.

[323] R Development Core Team, "R: a Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2015.

[324] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer International Publishing, 2016.

[325] S. Dudoit, Y. H. Yang, M. J. Callow, and T. P. Speed, "Statistical Methods for Identifying Differentially Expressed Genes in Replicated Microarray Experiments," *Stat. Sinca*, vol. 12, pp. 111–139, 2002.

[326] C. Feng *et al.*, "Log-transformation and its implications for data analysis," *Shanghai Arch. Psychiatry*, vol. 26, no. 2, pp. 105–109, 2014.

[327] I. T. Jollife, J. Cadima, I. T. Jolliffe, and J. Cadima, “Principal component analysis: A review and recent developments,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, p. 20150202, Apr. 2016.

[328] M. Ringnér, “What is principal component analysis?,” *Nat. Biotechnol.*, vol. 26, no. 3, pp. 303–304, Mar. 2008.

[329] J. Lever, M. Krzywinski, and N. Altman, “Principal component analysis,” *Nat. Methods*, vol. 14, 2017.

[330] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Second., vol. 27, no. 2. New York: Springer, 2009.

[331] M. R. Forster, “Key Concepts in Model Selection: Performance and Generalizability,” *J. Math. Psychol.*, vol. 231, pp. 205–231, 2000.

[332] J. C. Stoltzfus, “Logistic regression: A brief primer,” *Acad. Emerg. Med.*, vol. 18, no. 10, pp. 1099–1104, Oct. 2011.

[333] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 6th ed. Pearson, 2013.

[334] K. G. M. Moons *et al.*, “Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration,” *Ann. Intern. Med.*, vol. 162, no. 1, pp. W1–W73, Jan. 2015.

[335] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.

[336] J. Friedman, T. Hastie, R. Tibshirani, and N. Simon, “glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models.” 2019.

[337] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 68, no. 1, pp. 49–67, 2006.

[338] Y. Yang and H. Zou, “A fast unified algorithm for solving group-lasso penalize learning problems,” *Stat. Comput.*, vol. 25, no. 6, pp. 1129–1141, 2015.

[339] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification & Regression Trees*. Belmont, CA: Wadsworth Publishing Group, 1984.

[340] J. A. Hanley and B. J. McNeil, “The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve,” *Radiology*, vol. 143, no. 1, pp. 29–36, Apr. 1982.

[341] D. G. Altman, Y. Vergouwe, P. Royston, and K. G. M. Moons, “Prognosis and prognostic research: Validating a prognostic model,” *BMJ*, vol. 338, no. 7708, pp. 1432–1435, Jun. 2009.

[342] T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer, “ROCR: Visualizing the

performance of scoring classifiers.” 2013.

[343] J. B. Du Prel, B. Röhrig, G. Hommel, and M. Blettner, “Choosing Statistical Tests,” *Dtsch. Arztebl.*, vol. 107, no. 19, pp. 343–348, 2010.

[344] J. T. J. Huang *et al.*, “Independent protein-profiling studies show a decrease in apolipoprotein A1 levels in schizophrenia CSF, brain and peripheral tissues,” *Mol. Psychiatry*, vol. 13, no. 12, pp. 1118–1128, Dec. 2008.

[345] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders, fourth edition*. American Psychiatric Association, 2001.

[346] H. Akaike, “Information theory and an extension of the maximum likelihood principle,” in *Second International Symposium on Information Theory*, 1973, pp. 267–281.

[347] C. M. Hurvich and C. L. Tsai, “Regression and time series model selection in small samples,” *Biometrika*, vol. 76, no. 2, pp. 297–307, Jun. 1989.

[348] N. Sugiura, “Further analysis of the data by akaike’ s information criterion and the finite corrections,” *Commun. Stat. - Theory Methods*, vol. 7, no. 1, pp. 13–26, Jan. 1978.

[349] M. K. Chan *et al.*, “Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset,” *Transl. Psychiatry*, vol. 5, no. 7, p. e601, Jul. 2015.

[350] Y. Levin, L. Wang, E. Schwarz, D. Koethe, F. M. Leweke, and S. Bahn, “Global proteomic profiling reveals altered proteomic signature in schizophrenia serum,” *Mol. Psychiatry*, vol. 15, no. 11, pp. 1088–1100, Nov. 2010.

[351] Y. Yang *et al.*, “Altered levels of acute phase proteins in the plasma of patients with schizophrenia,” *Anal. Chem.*, vol. 78, no. 11, pp. 3571–3576, Jun. 2006.

[352] C. Wan *et al.*, “Abnormal changes of plasma acute phase proteins in schizophrenia and the relation between schizophrenia and haptoglobin (Hp) gene,” *Amino Acids*, vol. 32, no. 1, pp. 101–108, Jan. 2007.

[353] World Health Organization, “International statistical classification of diseases and related health problems - 10th revision,” *World Heal. Organ.*, vol. 2, no. 5, pp. 1–252, 2011.

[354] B. W. J. H. Penninx *et al.*, “The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods,” *Int. J. Methods Psychiatr. Res.*, vol. 17, no. 3, pp. 121–140, Sep. 2008.

[355] A. J. Rush, C. M. Gullion, M. R. Basco, R. B. Jarrett, and M. H. Trivedi, “The Inventory of Depressive Symptomatology (IDS): psychometric properties,” *Psychol. Med.*, vol. 26, no. 03, p. 477, May 1996.

[356] A. J. Rush *et al.*, "The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression," *Biol. Psychiatry*, vol. 54, no. 5, pp. 573–583, 2003.

[357] S. van Buuren and K. Groothuis-Oudshoorn, "Multivariate Imputation by Chained Equations," *J. Stat. Softw.*, vol. 45, no. 3, pp. 1–67, 2011.

[358] W. W. Eaton, M. Badawi, and B. Melton, "Prodromes and precursors: Epidemiologic data for primary prevention of disorders with slow onset," *Am. J. Psychiatry*, vol. 152, no. 7, pp. 967–972, Jul. 1995.

[359] The UniProt Consortium, "UniProt: a hub for protein information," *Nucleic Acids Res.*, 2015.

[360] E. A. Lannan, A. J. Galliher-Beckley, A. B. Scoltock, and J. A. Cidlowski, "Proinflammatory actions of glucocorticoids: Glucocorticoids and TNFα coregulate gene expression in vitro and in vivo," *Endocrinology*, vol. 153, no. 8, pp. 3701–3712, Aug. 2012.

[361] R. Troubat *et al.*, "Neuroinflammation and depression: A review," *Eur. J. Neurosci.*, vol. 53, no. 1, pp. 151–171, 2021.

[362] G. Fanelli *et al.*, "Reduced plasma Fetuin-A is a promising biomarker of depression in the elderly," *Eur. Arch. Psychiatry Clin. Neurosci.*, vol. 270, no. 7, pp. 901–910, Oct. 2020.

[363] S. Park, Y. Y. Nam, Y. Sim, and J. P. Hong, "Interactions between the apolipoprotein E ε4 allele status and adverse childhood experiences on depressive symptoms in older adults," *Eur. J. Psychotraumatol.*, vol. 6, no. 1, pp. 1–7, 2015.

[364] D. Gallagher-Thompson, R. Ohara, A. Simmons, H. C. Kraemer, and G. M. Murphy, "Apolipoprotein E ε4 allele affects the relationship between stress and depression in caregivers of patients with Alzheimer's disease," *J. Geriatr. Psychiatry Neurol.*, vol. 14, no. 3, pp. 115–119, Sep. 2001.

[365] M. Sadeghi *et al.*, "Relationship between depression and apolipoproteins A and B: a case-control study," *Clinics*, vol. 66, no. 1, pp. 113–117, 2011.

[366] S. Dockray, E. J. Susman, and L. D. Dorn, "Depression, Cortisol Reactivity, and Obesity in Childhood and Adolescence," *J. Adolesc. Heal.*, vol. 45, no. 4, pp. 344–350, Oct. 2009.

[367] A. Revah-Levy *et al.*, "Association between Body Mass Index and depression: The 'fat and jolly' hypothesis for adolescents girls," *BMC Public Health*, vol. 11, p. 649, 2011.

[368] T. Fryers, D. Melzer, and R. Jenkins, “Social inequalities and the common mental disorders: A systematic review of the evidence,” *Soc. Psychiatry Psychiatr. Epidemiol.*, vol. 38, no. 5, pp. 229–237, 2003.

[369] P. Bracke, E. Pattyn, and O. Von dem Knesebeck, “Overeducation and depressive symptoms: Diminishing mental health returns to education,” *Sociol. Heal. Illn.*, vol. 35, no. 8, pp. 1242–1259, 2013.

[370] I. Bjelland, S. Krokstad, A. Mykletun, A. A. Dahl, G. S. Tell, and K. Tambs, “Does a higher educational level protect against anxiety and depression? The HUNT study,” *Soc. Sci. Med.*, vol. 66, no. 6, pp. 1334–1345, 2008.

[371] L. Boschloo *et al.*, “The Network Structure of Symptoms of the Diagnostic and Statistical Manual of Mental Disorders,” *PLoS One*, vol. 10, no. 9, p. e0137621, Sep. 2015.

[372] E. I. Fried and R. M. Nesse, “Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study,” *J. Affect. Disord.*, vol. 172, pp. 96–102, Feb. 2014.

[373] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, “The PHQ-9: Validity of a brief depression severity measure,” *J. Gen. Intern. Med.*, vol. 16, no. 9, pp. 606–613, 2001.

[374] R. L. Spitzer, K. Kroenke, and J. B. W. Williams, “Validation and utility of a self-report version of PRIME-MD,” *Prim. Care Companion J. Clin. Psychiatry*, vol. 2, no. 1, p. 31, 2000.

[375] R. Tennant *et al.*, “The Warwick-Edinburgh mental well-being scale (WEMWBS): Development and UK validation,” *Health Qual. Life Outcomes*, vol. 5, no. 1, p. 63, Nov. 2007.

[376] P. Costa and R. R. McCrae, *NEO PI-R professional manual: Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa, FL, 1992.

[377] T. Välikangas, T. Suomi, and L. L. Elo, “A systematic evaluation of normalization methods in quantitative label-free proteomics,” *Brief. Bioinform.*, vol. 86, p. bbw095, Oct. 2016.

[378] P. Filzmoser, B. Liebmann, and K. Varmuza, “Repeated double cross validation,” *J. Chemom.*, vol. 23, no. 4, pp. 160–171, 2009.

[379] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, 1st ed. Chapman and Hall/CRC, 1993.

[380] Y. Benjamini and Y. Hochberg, “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing,” *J. R. Stat. Soc. Ser. B*, vol. 57, no. 1, pp. 289–

300, 1995.

[381] A. Rajagopalan, P. Shah, M. W. Zhang, and R. C. Ho, “Digital platforms in the assessment and monitoring of patients with bipolar disorder,” *Brain Sci.*, vol. 7, no. 11, 2017.

[382] M. Faurholt-Jepsen, M. Frost, E. M. Christensen, J. E. Bardram, M. Vinberg, and L. V. Kessing, “The effect of smartphone-based monitoring on illness activity in bipolar disorder: The MONARCA II randomized controlled single-blinded trial,” *Psychol. Med.*, vol. 50, no. 5, pp. 838–848, Apr. 2020.

[383] E. Nieto, E. Vieta, L. Alvarez, M. Torra, F. Colom, and C. Gastó, “Alpha-1-acid glycoprotein in major depressive disorder: Relationships to severity, response to treatment and imipramine plasma levels,” *J. Affect. Disord.*, vol. 59, no. 2, pp. 159–164, Aug. 2000.

[384] D. Healy *et al.*, “Alpha-1-acid glycoprotein in major depressive and eating disorders,” *J. Affect. Disord.*, vol. 22, no. 1–2, pp. 13–20, 1991.

[385] C. B. Nemeroff, K. R. Krishnan, D. G. Blazer, D. L. Knight, D. Benjamin, and L. R. Meyerson, “Elevated Plasma Concentrations of α1-Acid Glycoprotein, a Putative Endogenous Inhibitor of the Tritiated Imipramine Binding Site, in Depressed Patients,” *Arch. Gen. Psychiatry*, vol. 47, no. 4, pp. 337–340, Apr. 1990.

[386] Z. Kronfol and J. D. House, “Lymphocyte mitogenesis, immunoglobulin and complement levels in depressed patients and normal controls,” *Acta Psychiatr. Scand.*, vol. 80, no. 2, pp. 142–147, Aug. 1989.

[387] C. Song, T. Dinan, and B. E. Leonard, “Changes in immunoglobulin, complement and acute phase protein levels in the depressed patients and normal controls,” *J. Affect. Disord.*, vol. 30, no. 4, pp. 283–288, Apr. 1994.

[388] C. Shin *et al.*, “Increased plasma complement factor H is associated with geriatric depression,” *Int. Psychogeriatrics*, vol. 31, no. 1, pp. 1–8, May 2018.

[389] K. R. R. Krishnan *et al.*, “Apolipoprotein E-ε4 frequency in geriatric depression,” *Biol. Psychiatry*, vol. 40, no. 1, pp. 69–71, Jul. 1996.

[390] A. S. Rigaud *et al.*, “Association of the apolipoprotein E ε4 allele with late-onset depression,” *Neuroepidemiology*, vol. 20, no. 4, pp. 268–272, 2001.

[391] R. J. Caselli *et al.*, “Longitudinal changes in cognition and behavior in asymptomatic carriers of the APOE e4 allele,” *Neurology*, vol. 62, no. 11, pp. 1990–1995, Jun. 2004.

[392] J. D. Bremner, P. McCaffery, J. Douglas Bremner, and P. McCaffery, “The neurobiology of retinoic acid in affective disorders,” *Prog. Neuro-Psychopharmacology*

Biol. Psychiatry, vol. 32, no. 2, pp. 315–331, Feb. 2008.

[393] R. A. Hoff, M. Livingston Bruce, S. V. Kasl, and S. C. Jacobs, "Subjective ratings of emotional health as a risk factor for major depression in a community sample," *Br. J. Psychiatry*, vol. 170, no. FEB., pp. 167–172, 1997.

[394] J. A. Fleishman and S. H. Zuvekas, "Global self-rated mental health: Associations with other mental health measures and with role functioning," *Med. Care*, vol. 45, no. 7, pp. 602–609, Jul. 2007.

[395] S. H. Zuvekas and J. A. Fleishman, "Self-rated mental health and racial/ethnic disparities in mental health service use," *Med. Care*, vol. 46, no. 9, pp. 915–923, Sep. 2008.

[396] M. Olfson, S. C. Marcus, M. Tedeschi, and G. J. Wan, "Continuity of antidepressant treatment for adults with depression in the United States," *Am. J. Psychiatry*, vol. 163, no. 1, pp. 101–108, Jan. 2006.

[397] D. D. McAlpine, E. McCreedy, and S. Alang, "The Meaning and Predictive Value of Self-rated Mental Health among Persons with a Mental Health Problem," *J. Health Soc. Behav.*, vol. 59, no. 2, pp. 200–214, Jun. 2018.

[398] I. H. Benedek, R. A. Blouin, and P. J. McNamara, "Serum protein binding and the role of increased alpha 1-acid glycoprotein in moderately obese male subjects," *Br. J. Clin. Pharmacol.*, vol. 18, no. 6, pp. 941–946, 1984.

[399] I. H. Benedek, W. D. Fiske Iii, W. O. Griffen, R. M. Bell, R. A. Blouin, and P. J. Mcnamara, "Serum alpha 1-acid glycoprotein and the binding of drugs in obesity," *Br. J. Clin. Pharmacol.*, vol. 16, pp. 751–754, 1983.

[400] A. Shettima and M. M. Ibrahim, "Low levels of serum complement factor H is associated with increasing progression of bronchiectasis," *Biomed. Res. Ther.*, vol. 6, no. 7, pp. 3286–3292, Jul. 2019.

[401] S. L. T. Pek *et al.*, "Association of circulating proinflammatory marker, leucine-rich-α2-glycoprotein (LRG1), following metabolic/bariatric surgery," *Diabetes. Metab. Res. Rev.*, vol. 34, no. 7, p. e3029, Oct. 2018.

[402] D. N. Klein, R. Kotov, and S. J. Bufferd, "Personality and Depression: Explanatory Models and Review of the Evidence," *Annu. Rev. Clin. Psychol.*, vol. 7, no. 1, pp. 269–295, 2011.

[403] A. G. Chambers, A. J. Percy, D. B. Hardie, and C. H. Borchers, "Comparison of proteins in whole blood and dried blood spot samples by LC/MS/MS," *J. Am. Soc. Mass Spectrom.*, vol. 24, no. 9, pp. 1338–1345, Sep. 2013.

[404] N. J. Martin, J. Bunch, and H. J. Cooper, "Dried blood spot proteomics: Surface extraction of endogenous proteins coupled with automated sample preparation and mass spectrometry analysis," *J. Am. Soc. Mass Spectrom.*, vol. 24, no. 8, pp. 1242–1249, Aug. 2013.

[405] A. Steptoe, M. Hamer, and Y. Chida, "The effects of acute psychological stress on circulating inflammatory factors in humans: A review and meta-analysis," *Brain. Behav. Immun.*, vol. 21, no. 7, pp. 901–912, Oct. 2007.

[406] I. Shapira-Lichter *et al.*, "Cytokines and cholinergic signals co-modulate surgical stress-induced changes in mood and memory," *Brain. Behav. Immun.*, vol. 22, no. 3, pp. 388–398, Mar. 2008.

[407] L. de Wit, F. Luppino, A. van Straten, B. Penninx, F. Zitman, and P. Cuijpers, "Depression and obesity: A meta-analysis of community-based studies," *Psychiatry Res.*, vol. 178, no. 2, pp. 230–235, 2010.

[408] M. F. O'Connor *et al.*, "To assess, to control, to exclude: Effects of biobehavioral factors on circulating inflammatory markers," *Brain. Behav. Immun.*, vol. 23, no. 7, pp. 887–897, Oct. 2009.

[409] G. E. Miller, K. E. Freedland, R. M. Carney, C. A. Stetler, and W. A. Banks, "Pathways linking depression, adiposity, and inflammatory markers in healthy young adults," *Brain. Behav. Immun.*, vol. 17, no. 4, pp. 276–285, Aug. 2003.

[410] J. G. F. M. Hovens *et al.*, "Childhood life events and childhood trauma in adult patients with depressive, anxiety and comorbid disorders vs. controls," *Acta Psychiatr. Scand.*, vol. 122, no. 1, pp. 66–74, Oct. 2010.

[411] P. Rohde *et al.*, "Associations of child sexual and physical abuse with obesity and depression in middle-aged women," *Child Abus. Negl.*, vol. 32, no. 9, pp. 878–887, Sep. 2008.

[412] T. B. Gustafson and D. B. Sarwer, "Childhood sexual abuse and obesity," *Obes. Rev.*, vol. 5, no. 3, pp. 129–135, 2004.

[413] L. Davis, A. J. Barnes, A. C. Gross, J. R. Ryder, and R. J. Shlafer, "Adverse Childhood Experiences and Weight Status among Adolescents," *J. Pediatr.*, vol. 204, pp. 71–76, 2019.

[414] M. Macht, "How emotions affect eating: A five-way model," *Appetite*, vol. 50, no. 1, pp. 1–11, Jan. 2008.

[415] G. L. Palmisano, M. Innamorati, and J. Vanderlinden, "Life adverse experiences in relation with obesity and binge eating disorder: A systematic review," *J. Behav. Addict.*,

vol. 5, no. 1, pp. 11–16, 2016.

[416] C. Evers, F. Stok, and D. T. D. de Ridder, “Feeding your feelings: Emotion regulation strategies and emotional eating,” *Personal. Soc. Psychol. Bull.*, vol. 36, no. 6, pp. 792–804, Jun. 2010.

[417] V. Michopoulos, A. Powers, C. Moore, S. Villarreal, K. J. Ressler, and B. Bradley, “The mediating role of emotion dysregulation and depression on the relationship between childhood trauma exposure and emotional eating,” *Appetite*, vol. 91, pp. 129–136, Aug. 2015.

[418] M. Jokela, M. Virtanen, G. D. Batty, and M. Kivimäki, “Inflammation and Specific Symptoms of Depression,” *JAMA Psychiatry*, vol. 73, no. 1, p. 87, 2016.

[419] H. E. Duivis, N. Vogelzangs, N. Kupper, P. De Jonge, and B. W. J. H. Penninx, “Differential association of somatic and cognitive symptoms of depression and anxiety with inflammation: Findings from the netherlands study of depression and anxiety (NESDA),” *Psychoneuroendocrinology*, vol. 38, no. 9, pp. 1573–1585, Sep. 2013.

[420] A. L. Chu, J. Stochl, G. Lewis, S. Zammit, P. B. Jones, and G. M. Khandaker, “Longitudinal association between inflammatory markers and specific symptoms of depression in a prospective birth cohort,” *Brain. Behav. Immun.*, vol. 76, pp. 74–81, Feb. 2019.

[421] R. Dantzer, J. C. O’Connor, G. G. Freund, R. W. Johnson, and K. W. Kelley, “From inflammation to sickness and depression: When the immune system subjugates the brain,” *Nat. Rev. Neurosci.*, vol. 9, no. 1, pp. 46–56, Jan. 2008.

[422] E. Iob, C. Kirschbaum, and A. Steptoe, “Persistent depressive symptoms, HPA-axis hyperactivity, and inflammation: the role of cognitive-affective and somatic symptoms,” *Mol. Psychiatry*, vol. 25, no. 5, pp. 1130–1140, 2020.

[423] E. I. Fried and R. M. Nesse, “Depression sum-scores don’t add up: why analyzing specific depression symptoms is essential,” *BMC Med.*, vol. 13, no. 1, p. 72, Dec. 2015.

[424] S. M. Monroe and A. D. Simons, “Diathesis-stress theories in the context of life stress research: implications for the depressive disorders,” *Psychol. Bull.*, vol. 110, no. 3, pp. 406–425, 1991.

[425] A. Arnau-Soler *et al.*, “A validation of the diathesis-stress model for depression in Generation Scotland,” *Transl. Psychiatry*, vol. 9, no. 1, 2019.

[426] B. G. Shapero *et al.*, “Stressful Life Events and Depression Symptoms: The Effect of Childhood Emotional Abuse on Stress Reactivity,” *J. Clin. Psychol.*, vol. 70, no. 3, pp. 209–223, Mar. 2014.

[427] K. S. Kendler *et al.*, “Stressful life events, genetic liability, and onset of an episode of major depression in women,” *Am. J. Psychiatry*, vol. 152, no. 6, pp. 833–842, Apr. 1995.

[428] K. S. Kendler, L. M. Karkowski, and C. A. Prescott, “Causal relationship between stressful life events and the onset of major depression,” *Am. J. Psychiatry*, vol. 156, no. 6, pp. 837–841, 1999.

[429] L. Colodro-Conde *et al.*, “A direct test of the diathesis–stress model for depression,” *Mol. Psychiatry*, vol. 23, no. 7, pp. 1590–1596, 2018.

[430] S. B. Patten, “Major depression epidemiology from a diathesis-stress conceptualization,” *BMC Psychiatry*, vol. 13, p. 19, 2013.

[431] S. R. S. Ouakinin, D. P. Barreira, and C. J. Gois, “Depression and Obesity: Integrating the Role of Stress, Neuroendocrine Dysfunction and Inflammatory Pathways,” *Front. Endocrinol. (Lausanne).*, vol. 9, p. 431, 2018.

[432] W. W. Wang, X. L. Liu, Y. Ruan, L. Wang, and T. H. Bao, “Depression was associated with apolipoprotein E ε4 allele polymorphism: A meta-analysis,” *Iran. J. Basic Med. Sci.*, vol. 22, no. 2, pp. 112–117, Feb. 2019.

[433] J. M. Ramsey *et al.*, “Sex differences in serum markers of major depressive disorder in the Netherlands Study of Depression and Anxiety (NESDA),” *PLoS One*, vol. 11, no. 5, 2016.

[434] M. C. Jentsch *et al.*, “Gender differences in developing biomarker-based major depressive disorder diagnostics,” *Int. J. Mol. Sci.*, vol. 21, no. 9, 2020.

[435] T. F. van de Mortel, “Faking it: Social desirability response bias in selfreport research,” *Aust. J. Adv. Nurs.*, vol. 25, no. 4, pp. 40–48, 2008.

[436] W. H. McDonald and J. R. Yates, “Shotgun proteomics and biomarker discovery,” *Dis. Markers*, vol. 18, no. 2, pp. 99–105, 2002.

[437] D. Allen and B. McWhinney, “Quadrupole Time-of-Flight Mass Spectrometry: A Paradigm Shift in Toxicology Screening Applications,” *Clin. Biochem. Rev.*, vol. 40, no. 3, pp. 135–146, Aug. 2019.