Academic Year 2019-2020

Faculty Pharmaceutical, Biomedical and Veterinary Sciences

Biomedical Sciences

Breaking the habit – Exploring the potential of peripheral blood mononuclear cells in comparison with plasma as a sample source for biomarker discovery in psychiatry

By:

Jolien Hendrix

Master Thesis in partial fulfilment of the requirements for the degree

Master in Biomedical Sciences

Promotor(s): Prof. Dr. Violette Coppens

Copromotor: Prof. Dr. Manuel Morrens

Coach: Jobbe Goossens

Scientific Initiative for Neuropsychiatric and Psychopharmacological Studies
Psychiatric Hospital Duffel
Rooienberg 21
2570 Duffel
Belgium

Table of Contents

Abstract

Introduction: The quality of psychiatric healthcare is significantly hampered by high rates of misdiagnosis, causing major psychiatric disorders to be among the leading causes of global disease burden. These high rates of misdiagnosis are partially a result of the extensive symptom overlap across major psychiatric disorders, more specifically major depressive disorder, bipolar disorder and schizophrenia. Furthermore, psychiatric diagnostics lack objective, biological measures to aid in the differential diagnosis of disorders with similar clinical presentations and is therefore affected by a high degree of subjectivity. In the last decades, many studies aimed to identify diagnostic biomarkers, mainly in plasma or serum. However, current experimental set-ups failed to yield any reliable biomarkers. Hence, the main goal of this thesis was to establish a robust mass spectrometry protocol for biomarker discovery by determining the best sample source, protein depletion strategy and analytical algorithm. Moreover, this thesis involved data analysis as part of a larger ongoing project of the host laboratory, which aims to identify biomarker candidates for differential diagnosis of the previously mentioned major psychiatric disorders.

Methodology: *Original experimental set-up*: Plasma and peripheral blood mononuclear cells of healthy volunteers (N = 2) were analysed via label-free liquid chromatography – mass spectrometry. To establish the optimal biomarker discovery protocol, 6 liquid chromatography elution gradient durations (155, 180, 240, 300, 360 and 440 minutes), 2 depletion columns (ProteoSpin™ Abundant Serum Protein Depletion Kit vs. Pierce™ Top 12 Abundant Protein Depletion Spin Column), 2 analytical algorithms (MaxQuant vs. MaxLFQ) and 2 sample sources (peripheral blood mononuclear cells vs. plasma) were compared for yield (i.e. the number of protein identifications) and protein retrieval reliability on multiple levels.
Replacement assignment: Biomarker candidates for differential diagnosis of major depressive disorder (N = 5) and the depressive state of bipolar disorder (N = 3), on the one hand, and the manic state of bipolar disorder (N = 4) and schizophrenia (N = 4), on the other, were identified in peripheral blood mononuclear cells using iTRAQ liquid chromatography – mass spectrometry. Age and gender matched healthy controls (N = 6) were included as baseline comparator. Subsequently, biomarker candidates were analysed via pathway (CAMERA) and network (Cytoscape) analysis.

Results: *Original experimental set-up*: The number of identifications was higher and overall variation was lower in plasma depleted with the ProteoSpin™ column than in plasma depleted with the Pierce™ column. In peripheral blood mononuclear cells, more proteins were identified and overall variation was higher than in plasma. Nevertheless, the number of proteins that were quantified with an acceptable precision was higher in peripheral blood mononuclear cells than in plasma. Analysis of mass spectra using MaxLFQ decreased the number of identifications and overall variation in plasma and peripheral blood mononuclear cells compared to analysis with the standard MaxQuant algorithm.
Replacement assignment: When comparing expression values of patients suffering from major depressive disorder and bipolar patients in a depressive state, 67 biomarker candidates for differential diagnosis were identified, 139 gene sets were significantly enriched and biomarker candidates showed little interaction. Seventy eight biomarker candidates were identified for differential diagnosis of the manic state of bipolar disorder and schizophrenia. No gene sets were significantly enriched in this case and the majority of biomarker candidates were connected to each other.

Conclusion: *Original experimental set-up*: Data were not sufficient to draw solid conclusions from. However, based on the limited data, liquid chromatography – mass spectrometry of peripheral blood mononuclear cells while using MaxLFQ seems to be the more reliable approach for untargeted biomarker discovery for psychiatric afflictions.
Replacement assignment: Several potential biomarker candidates were discovered for discriminatory diagnostics in the abovementioned patient groups. As the sample size was small, results should be validated in larger cohorts to identify true biological differences.

Dutch summary

Inleiding: De kwaliteit van de psychiatrische gezondheidszorg wordt aanzienlijk belemmerd door het hoge aantal misdiagnoses, waardoor psychiatrische stoornissen behoren tot de hoofdoorzaken van ziektelast. Misdiagnoses zijn deels te wijten aan de sterke symptoom overlap, vooral tussen depressieve stoornis, bipolaire stoornis en schizofrenie. Bovendien zijn er in de psychiatrische diagnostiek geen biologische hulpmiddelen voor de differentiële diagnose van ziektes met een vergelijkbare klinische presentatie. In de afgelopen decennia hebben vele studies getracht diagnostische biomerkers te identificeren, voornamelijk in plasma of serum. Deze experimentele opstellingen leverden echter geen betrouwbare biomerkers op. Het hoofddoel van deze thesis was dan ook om een massaspectrometrie protocol op te stellen door het bepalen van het beste staal, de beste depletie strategie en het beste analytisch algoritme. Bovendien omvatte deze thesis data analyse als onderdeel van een lopend project van het gastlaboratorium dat gericht is op het identificeren van biomerker kandidaten voor de differentiële diagnose van de eerder genoemde psychiatrische stoornissen.

Methodologie: *Oorspronkelijke experimentele opzet*: Plasma en perifere mononucleaire cellen van gezonde vrijwilligers (N = 2) werden geanalyseerd via chromatografie - massaspectrometrie. Om het optimale protocol voor biomerker identificatie tot stand te brengen, werden 6 chromatografie gradiënten (155, 180, 240, 300, 360 en 440 minuten), 2 depletiekolommen (ProteoSpin ™ Depletion Kit vs. Pierce™ Depletion Column), 2 analytische algoritmen (MaxQuant vs. MaxLFQ) en 2 stalen (perifere mononucleaire cellen vs. plasma) vergeleken op basis van het aantal identificaties en de variabiliteit op proteïne intensiteiten. *Vervangopdracht*: Biomerker kandidaten voor differentiële diagnose van depressieve stoornis (N = 5) en de depressieve toestand van bipolaire stoornis (N = 3), enerzijds, en de manische toestand van bipolaire stoornis (N = 4) en schizofrenie (N = 4), anderzijds, werden geïdentificeerd in perifere mononucleaire cellen met behulp van iTRAQ chromatografie - massaspectrometrie. Gezonde controles (N = 6) werden ook geanalyseerd. Vervolgens werden de biomerker kandidaten geanalyseerd via *pathway* (CAMERA) en netwerk (Cytoscape) analyse.

Resultaten: *Oorspronkelijke experimentele opzet*: Het aantal identificaties was hoger en de totale variatie was lager in plasma verwerkt met de ProteoSpin ™ kolom dan in plasma verwerkt met de Pierce ™ kolom. Meer proteïnes werden geïdentificeerd en de totale variatie was hoger in perifere mononucleaire cellen dan in plasma. Desondanks was het aantal eiwitten dat met een aanvaardbare variatie werd gekwantificeerd hoger in perifere mononucleaire cellen dan in plasma. MaxLFQ verminderde het aantal identificaties en de totale variatie in plasma en perifere mononucleaire cellen. *Vervangopdracht*: Bij het vergelijken van expressiewaarden van patiënten die lijden aan een depressieve stoornis en bipolaire patiënten in een depressieve toestand, werden 67 biomerker kandidaten geïdentificeerd, 139 genensets waren differentieel gereguleerd en biomerker kandidaten vertoonden weinig interactie. 78 biomerker kandidaten werden geïdentificeerd voor de differentiële diagnose van de manische toestand van bipolaire stoornis en schizofrenie. Er werden geen genensets geïdentificeerd en de meeste biomerker kandidaten waren met elkaar verbonden.

Conclusie: *Originele experimentele opstelling*: De gegevens waren niet voldoende om er betrouwbare conclusies uit te trekken. Echter lijken perifere mononucleaire cellen en MaxLFQ de betere optie voor de ontdekking van biomerkers. *Vervangopdracht*: Verschillende potentiële biomerker kandidaten werden ontdekt voor discriminerende diagnostiek van de bovengenoemde patiëntengroepen. Aangezien de gegevens werden verkregen uit een klein cohort, moeten expressiewaarden van de proteïnes worden nagegaan in grotere cohorten om echte biologische verschillen te identificeren.

Overview of the Master's Thesis: Before and After COVID-19

Original experimental set-up: The original experimental set-up consisted of three phases of which the first two made use of technical samples whereas samples of psychiatric patient would have been used for the latter. First, the protocol for sample preparation of plasma for liquid chromatography – mass spectrometry (LCMS) would have been optimised. An important step in this protocol is the depletion of highly abundant proteins. Therefore, this thesis would have compared two commercially available depletion columns. Although the columns were used to prep three plasma samples each, these data are not sufficient to provide a solid answer to the question which column performs best. In case COVID-19 would not have happened, the columns would have been used to prep more samples so that the power of the comparison would have been larger. Moreover, the effect of the length of the liquid chromatography (LC) elution gradient on the number of proteins that can be retrieved from samples would have been assessed in plasma and peripheral blood mononuclear cells (PBMCs). The optimal gradient length would have been determined for both sample sources before proceeding to the next phase. At the very last day students were allowed to enter the laboratory, a range of gradient lengths were tested on plasma samples. However, only one data point per gradient length could be collected. This parameter could not be varied for PBMCs. Second, data obtained from plasma and PBMCs would have been compared to decide which sample source provides the best characteristics for biomarker discovery in terms of quantitative and qualitative endpoints (i.e. highest number of identifications and lowest variation on protein intensities). Plasma samples would have been prepped using the optimised protocol while PBMCs would have been prepped using the protocol for sample preparation of cells that was already available at the host laboratory. Due to the circumstances, these samples could not be analysed simultaneously, as would have been the case in the original experimental set-up to limit time effects. Thus, the protein expression levels of plasma samples that were prepped with the best depletion column and obtained during the optimisation phase were directly compared with the results obtained from PBMCs, even though the samples were analysed with more than one month in between. Due to the limited time spend in the laboratory, not all previously anticipated levels of variation could be assessed. For instance, intra-individual variability would have been assessed by comparing samples of the same individuals collected at different time points. Moreover, the biological variation on protein abundances detected in these sample sources would have been determined. In a last phase, protein expression profiles of four patient groups would have been analysed via LCMS to identify biomarker candidates for differential diagnosis. This part of the thesis would have used plasma and PBMCs of healthy controls (HCs), patients with major depressive disorder (MDD), bipolar patients in a depressive state (BD-D), bipolar patients in a manic state (BD-M) and schizophrenia patients with active psychotic symptoms (SZ). Subsequently, these biomarker candidates would have been analysed via various bioinformatic approaches. This final phase of the original experimental set-up could not be performed.

Replacement assignment: The biomarker discovery of the original experimental set-up was replaced by the analysis of data that were already available in the host laboratory. These data were obtained from PBMCs of the same experimental groups that would have been used in the original experimental set-up. This way, the most important part of this thesis, namely biomarker discovery for differential diagnosis of major psychiatric disorders, was not lost. Data analysis was performed in the same manner as would have been the case in the original experimental set-up. Hence, biomarker candidates for differential diagnosis of major psychiatric disorders were identified and analysed via pathway and network analysis.

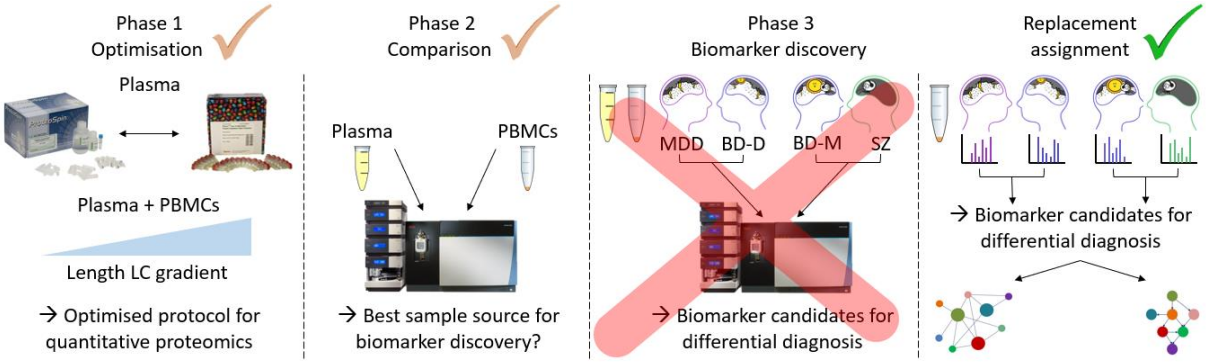

Figure 0 Overview of the master's thesis before and after COVID-19.
Abbreviations: BD-D, bipolar patients in a depressive state; BD-M, bipolar patients in a manic state; LC, liquid chromatography; MDD, patients with major depressive disorder; PBMCs, peripheral blood mononuclear cells; SZ, schizophrenia patients with active psychotic symptoms

1. Introduction

According to the latest Global Burden of Disease (GBD) Study[1], major depressive disorder (MDD), bipolar disorder (BD) and schizophrenia (SZ) are among the leading causes of global disease burden [1]. These morbidities belong to the major psychiatric disorders and present with a wide range of symptoms, varying from depressive episodes to impaired physical functions, such as sleep and appetite. All three disorders are hallmarked by cardinal symptoms, being depressive mood for MDD, extreme mood cycling for BD and psychotic symptoms for SZ [2-4]. However, this does not exclude the existence of these symptoms in other psychiatric disorders or imply that these symptoms represent the disorder's most frequent clinical presentation [5]. As shown in Figure 1, there is an extensive overlap of symptoms across these major psychiatric disorders, impeding proper diagnosis.

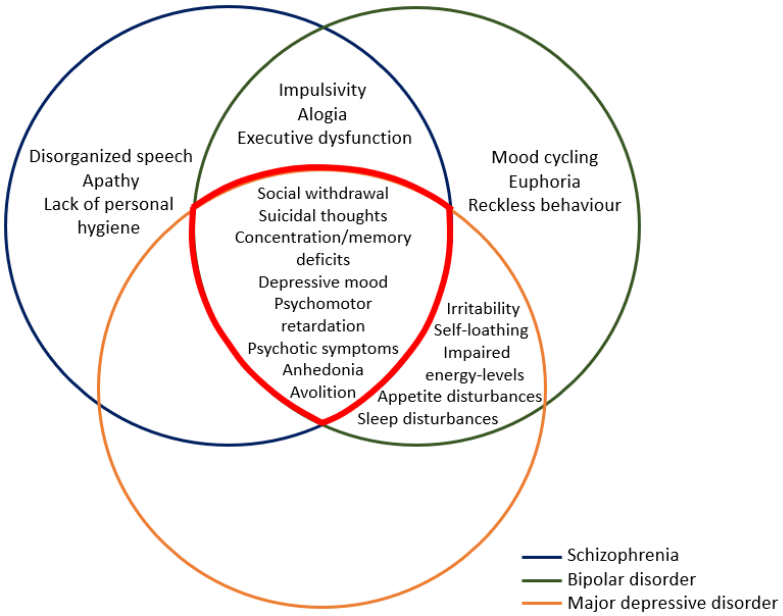

Figure 1 Symptoms of major depressive disorder, bipolar disorder and schizophrenia.

1.1 Burden of Major Psychiatric Disorders

1.1.1 Mortality

Annually, approximately 800.000 people die by suicide [6]. A vast majority of these deaths (90-95%) occurs in people with a diagnosable psychiatric disorder, as revealed by a psychological autopsy of suicide victims [7]. Moreover, psychiatric patients are up to 20 times more likely to commit suicide compared to the general population [8]. However, the latest GBD study reported that psychiatric disorders are only causing 0,0011% of total Years of Life Lost (YLLs), which is very little compared to the 20% of total YLLs that are caused by cardiovascular diseases for example. Even more astonishing is that the study reported that all mental illness-related YLLs are caused by eating disorders without mentioning separate numbers for MDD, BD or SZ and without linking YLLs caused by self-harm to these afflictions [9]. Of note, calculating YLLs caused by psychiatric disorders is extremely challenging as psychiatric disorders are often not the direct cause of death [10, 11].

Patients suffering from psychiatric disorders have an increased mortality rate compared to the general population, a discovery that was already made in 1937 and confirmed by more recent studies [8, 12].

[1] The Global Burden of Disease study is a comprehensive study defining global disease burden by assessing mortality and disability.

Specifically, the relative mortality risks of major psychiatric disorders are 1,6 for MDD, 2,2 for BD and 2,5 for SZ compared to the general population [8]. These high relative mortality risks can be explained by the increased prevalence of modifiable risk factors for all-cause mortality, such as smoking, poor diet and social deprivation, in psychiatric patients [13]. Consequently, such risk factors can result in the emergence of physical comorbidities associated with high mortality rates in the general population [14, 15]. When psychiatric patients die, these physical comorbidities are mostly considered to be the cause of death, even though the psychiatric condition drove those patients towards adverse behaviours leading to such morbidities [10, 11]. Moreover, the occurrence of physical comorbidities in and the prevalence of suicide among psychiatric patients contribute to a reduced life expectancy of 7 to 11 years for MDD patients, 9 to 20 years for BD patients and 10 to 20 years for SZ patients compared to the general population [8].

Furthermore, numerous studies report that the relative mortality risks of major psychiatric disorders increased over the last decades. This is further supported by the fact that the mortality gap between people with such disorders and the general population also increased, suggesting that people suffering from a major psychiatric disorder may not experience the increased life expectancy to the same extent as the general population [16-19].

1.1.2 Disability

Although many deaths are caused by psychiatric disorders, they only amass to a small percentage of the previously mentioned YLLs, which was the standard measure to assess disease burden in early studies [9, 20]. However, disease burden associated with major psychiatric disorders is mainly rooted in impaired daily functioning of patients and thus disability rather than mortality, as illustrated in Table 1. Therefore, the disease burden caused by major psychiatric disorders was seriously underestimated up until the first GBD study that was performed in 1990. The true burden of these afflictions was highlighted by the introduction of a new measure to quantify disease burden, namely the Disability-Adjusted Life Year (DALY). A DALY is the summation of YLLs and Years Lost due to Disability (YLDs), thereby capturing both mortality and disability. Simply said, one DALY represents one lost year of healthy life, either due to premature death or disability [20].

The latest GBD study showed that disease burden caused by psychiatric disorders accounted for nearly 5% of total DALY counts worldwide. As approximately half of these DALYs was caused by MDD, BD and SZ, these afflictions are considered to be the main drivers of disease burden associated with psychiatric disorders. Although this percentage might seem rather limited, psychiatric disorders are the sixth leading cause of DALYs and are therefore major contributors to global disease burden [21]. Compared to the percentage of DALY counts caused by psychiatric disorders, the YLD counts associated with these disorders were considerably larger (nearly 15% of total YLDs) [1]. Moreover, this difference between DALY and YLD counts, illustrated in Table 1, confirms that psychiatric disorders mainly contribute to disease burden by causing disability and not mortality, which is also illustrated by their rank in the list of leading causes of DALY and YLD counts. As previously mentioned, psychiatric disorders are the sixth leading cause of DALYs whereas they are the second leading cause of YLDs [1, 21].

As predicted by the GBD 1990 study and illustrated in Figure 2, the latest GBD study showed a significant increase in DALY counts for major psychiatric disorders, implying that psychiatric disorders are a growing issue for global health [21, 22][2]. Moreover, the World Health Organization (WHO) predicted in 2010 that psychiatric disorders will become the largest contributors to global disease burden by 2030 [23].

[2] This statement is arguable as the increase in DALY counts may also be a result of the increased awareness and visibility of psychiatric disorders, especially as these disorders are among the most stigmatised ones. Therefore, extrapolating might be challenging.

Table 1 Percentages of total YLLs, YLDs and DALYs caused by mental disorders and cardiovascular diseases in 2017.

Disorder(s)	Percentage of total YLLs	Percentage of total YLDs	Percentage of total DALYs
Major depressive disorder	*	3,86%	1,31%
Bipolar disorder	*	1,09%	0,37%
Schizophrenia	*	1,5%	0,51%
All mental disorders	0,0011%	14,41%	4,89%
Cardiovascular diseases	20,06%	4,19%	14,66%

Legend: *, Data not available
Abbreviations: DALYs, Disability-Adjusted Life Years; YLDs, Years Lost due to Disability; YLLs, Years of Life Lost

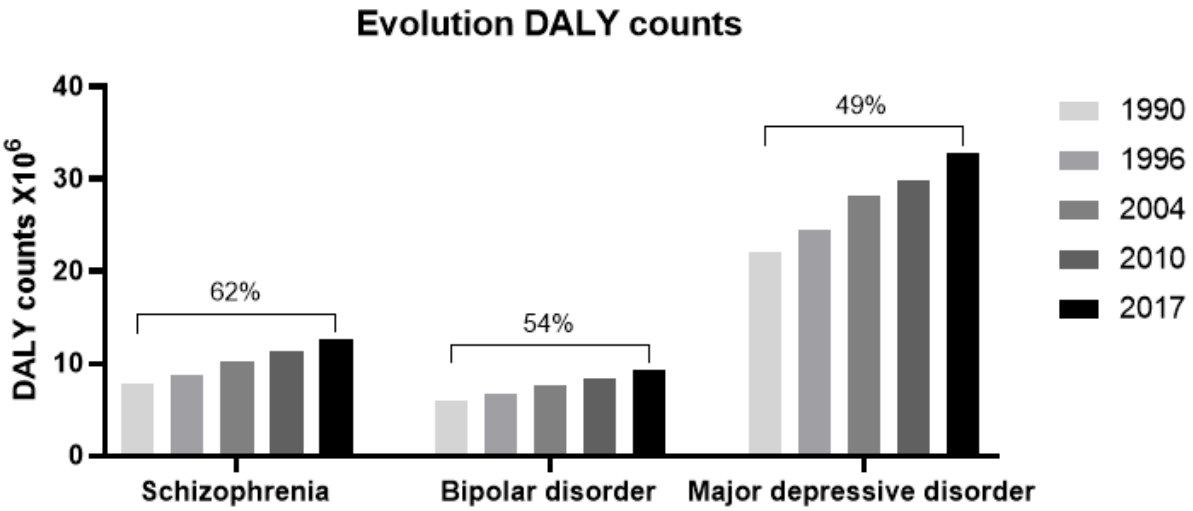

Figure 2 Evolution of DALY counts between 1990 and 2017. The percentages indicate the magnitude of the increases in DALY counts caused by schizophrenia, bipolar disorder and major depressive disorder. DALY counts are based on data of the GBD studies performed throughout the years.
Abbreviations: DALY, Disability-Adjusted Life Year

1.1.3 Economic Burden

Next to the impact of a psychiatric disorder on an individual, they also have an enormous impact on the economy. In 2011, the World Economic Forum published the global economic burden of non-communicable diseases, which revealed concerning numbers. The results showed that the expenditure associated with psychiatric disorders, costing the global economy annually US$ 2,5 trillion, was higher than that of chronic somatic diseases, such as cancer and diabetes. Based on prediction models, these costs are expected to increase significantly in the future [24].

The economic burden of diseases is determined by evaluating direct and indirect costs. While direct costs cover the actual expenditures, such as medical costs, indirect costs are described by the monetary value of lost resources caused by the disease, such as the loss of income due to reduced productivity of patients and caregivers [25]. Contrasting to other diseases, such as cardiovascular diseases and cancer, the economic burden of psychiatric disorders is dominated by indirect costs, which account for two-thirds of the total cost [24, 26, 27].

By accounting for 60% to even up to 90% of indirect costs and nearly 40% of total costs, the loss of income of patients and caregivers is the main driver of economic burden caused by psychiatric disorders [26, 27]. This is further supported by the fact that employees suffering from a psychiatric

disorder are on average more days out of role than healthy ones. Data from the World Mental Health (WMH) surveys[3] revealed that, annually, individuals with a psychiatric disorder were 31,1 more days out of role compared to those with no conditions. More specifically, employees diagnosed with MDD, BD or SZ were respectively 34,4, 41,2 and 40,3 more days out of role in one year [26, 28]. These numbers exceed the mean days out of role for individuals suffering from cancer (31,9) or a cardiovascular disease (28,7) [28] and confirm the results of the GBD studies, highlighting the severe disability and therefore also loss of productivity caused by these psychiatric disorders.

1.2 Diagnosis of Psychiatric Disorders

1.2.1 How Psychiatric Disorders should be Diagnosed

Following guidelines, diagnosis of psychiatric disorder solely depends on patient anamnesis which is referenced to diagnostic manuals. They are considered to be classification systems and list criteria associated with a certain disorder. Patients should meet at least several of the listed criteria to be diagnosed with a certain disorder [29]. Hence, psychiatric disorders are mainly diagnosed by confirming cardinal symptoms and eliminating most probable differentials [30].

The WHO as well as the American Psychiatric Association (APA) developed manuals to diagnose psychiatric disorders in a structured manner. The WHO's International Statistical Classification of Diseases and Related Health Problems, 10th Edition (ICD-10) manual is used for clinical purposes worldwide, with the exception of the United States [31]. The APA's Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) manual is used in clinical settings in the United States and for research purposes internationally [32].

1.2.2 How Psychiatric Disorders are actually Diagnosed in Practice

Although diagnostic manuals are available to apply a structured approach during the diagnostic process, unstructured interviews are still commonly used in practice. During such unstructured interviews, heuristic techniques, specifically the prototype and theory-based approach, are regularly applied [33]. The former approach is based on similarity-matching where a patient is compared to a prototype case of a disorder [34], while physicians using the latter conceive their own theories about a disorder instead of using manuals to diagnose psychiatric disorders [35]. Following Schneider's "first-rank symptoms" for example, (auditory) hallucinations have been suggested to be of high diagnostic value for SZ [36]. Therefore, a prototype case of SZ presents with psychotic symptoms while prototype cases of MDD and BD do not. However, many studies reported the existence of psychotic symptoms in MDD and BD patients with a lifetime prevalence of 28% for MDD and even up to 90% for BD patients [37-41] (for review see [42]).

Both heuristic approaches allow that the patient's diagnosis is influenced by the physician's experiences and personal views [33], which may be affected by selective memory and thus may result in systematic errors [35]. Consequently, the diagnostic process is highly subjective, leading to a decreased accuracy when using an unstructured approach compared to a structured approach and ultimately misdiagnosis of psychiatric disorders [43].

[3] The WMH surveys are part of the WMH initiative launched by the World Health Organization (WHO). It aims to obtain accurate cross-national information about the prevalence and correlate of mental, substance, and behavioural disorders.

1.2.3 Limitations of the Current Diagnostic Process

1.2.3.1 Limitations related to the Diagnostic Manuals

Various studies proved that current diagnostic tools fail to identify discrete psychiatric disorders and are, therefore, considered to be scientifically meaningless [5]. One limitation of the diagnostic manuals concerns the use of different criteria in a category representing the same psychiatric disorder across different manuals [44-46]. Hence, a patient might be diagnosed with one disorder when using DSM-5 while being diagnosed with another when using ICD-10. However, even when implementing a structured approach using the same manual, diagnostic reliabilities remain suboptimal. A recent meta-analysis assessed the interrater reliability in psychiatric disorders and reported a Cohen's kappa of 0,80 for SZ, 0,82 for BD and 0,75 for MDD. Moreover, the meta-analysis did not show any association between kappa values and publication year, implying that the interrater reliability in psychiatric disorders did not improve over the covered time period (1974-2012) [47]. In another meta-analysis, the same authors determined the Cohen's kappa values to evaluate the test-retest reliability in psychiatric disorders. These kappa values averaged at 0,7, which is slightly lower when compared to those representing the interrater reliability, and were negatively associated with the time between baseline and follow-up diagnosis for all three afflictions [48]. Therefore, this meta-analysis revealed that a psychiatric diagnosis is dynamic over time in a considerable proportion of the patient population. Although the cause of this remains unclear, an initial misdiagnosis may partially explain this phenomenon.

Additionally, diagnostic manuals are classification systems because a patient should meet a minimum (defined per disorder) number of the listed criteria, allowing that two patients can be diagnosed with the same disorder without having any symptoms or characteristics in common. Therefore, the diagnostic manuals represent psychiatric disorders as disjunctive categories. In 1968, Bannister was the first to address this problem for schizophrenia [49], but the same holds true for other psychiatric disorders, such as MDD and BD [5]. Although both DSM and ICD manuals have been adjusted over the years, psychiatric disorders are still represented by disjunctive categories [50]. Furthermore, these categories do not consider the existence of the same symptoms across psychiatric disorders and thus, patients do not always fit within the boundaries of a single diagnostic category, even though this may also be a consequence of the nature of the disorders [5]. This is especially concerning knowing that some symptoms are only included in the list of criteria for a certain disorder while they are also prevalent in others. Illustratively, the category representing SZ in the ICD-10 manual includes some of Schneider's "first-rank symptoms", which may result in misdiagnosis when such symptoms are present in patients suffering from another psychiatric disorder [51].

Moreover, the criteria listed in the diagnostic manuals are strict in terms of duration and amount or severity of symptoms to receive a certain diagnosis. Consequently, individuals presenting with subthreshold symptoms may remain undiagnosed even though they should be treated following the guidelines for that disorder [52]. Furthermore, the diagnostic manuals allow the use of different questionnaires to obtain the necessary information as it is not specified which one should be used. This increases the subjectivity involved in the diagnostic process because each user of the diagnostic manuals decides which questionnaire is best suited for themselves and thus also the heterogeneity within the diagnostic process [53].

1.2.3.2 Limitations related to the Nature of the Disorders

As the symptomatology of psychiatric patients is broad and varies with time, the clinical presentation of such patients is a dynamic entity. Hence, a major issue with diagnosing such patients concerns the time point at which they consult the physician since there is a possibility that not all previously experienced symptoms are present at that specific moment. When the patient does not recall certain

symptoms, important characteristics of a disorder may be overseen [44]. Moreover, some symptoms might not be recognised as pathological, which also results in a failure to report these symptoms [54, 55]. For example, (hypo-)manic characteristics, such as elevated energy and euphoria, are important symptoms to mention during a clinical interview since they are cardinal symptoms of BD [2]. However, it can be challenging for patients to recognise those symptoms as being abnormal, resulting in a failure to report them. Watanabe et al. showed that 39% of BD patients did not recognise (hypo-)manic symptoms and thus failed to report them to the physician [56]. Although diagnostic guidelines require physicians to probe for symptoms of other disorders to eliminate the most probable differentials, the use of heuristic approaches sustains the impact of the time point of consultation [31, 32]. Combined with the fact that BD patients frequently consult a physician in a depressed state, this might cloud the physician's judgement [57].

1.2.4 High Rates of Misdiagnosis

The underlying key issue of the abovementioned limitations is the lack of objective measurements, which could confirm or refute the physician's diagnosis. The extensive overlap of symptoms across MDD, BD and SZ enlarges this problem of subjectivity as some symptoms may be valued more than others based on a prototype case of or the physician's personal theory about a certain disorder [35]. As is to be expected, this *modus operandi* of the current diagnostic process results in alarmingly high rates of misdiagnosis in psychiatry.

Frightening numbers implicate that especially the diagnosis of BD seems to be challenging. Only 1 in 5 BD patients is correctly diagnosed within 1 year after consulting a physician, whereas approximately 70% received at least one misdiagnosis [56, 58, 59]. On average, it takes 3,5 misdiagnoses and 4 physicians over a time period of 5 to 10 years prior to receiving the correct diagnosis of BD [58, 60, 61]. This time period even exceeded 10 years in one third of BD patients examined by Hirschfield et al. [58].

With an initial MDD diagnosis in 65% of misdiagnosed BD patients, MDD is the most frequent misdiagnosis of BD [56, 58, 59]. This may be explained by the high prevalence of depressive symptoms and low prevalence of typical (hypo-)manic symptoms in BD patients. Judd et al. observed BD patients over a time period of 13 years and concluded that BD patients were in a (hypo-)manic state less than 10% of the time, whereas patients reported to be depressed nearly 40% of the time [62]. Furthermore, the clinical presentation of BD patients presenting to the physician with solely depressive symptoms is very similar to the presentation of MDD patients [63]. This is supported by results obtained by Watanabe and colleagues, who showed that 84% of BD patients without an early correct diagnosis[4] presented in a depressive state at the first visit compared to 43% of BD with an early correct diagnosis [56].

As previously mentioned, psychotic symptoms are suggested to be of high diagnostic value for SZ and are therefore frequently implemented in heuristic approaches and even the ICD-10 diagnostic manual to reach differential diagnosis [51]. Consequently, this leads to BD patients receiving the diagnosis of SZ since psychotic symptoms are common in this population [37-40, 64]. More than one in five BD patients presenting with psychotic symptoms are misdiagnosed with a subtype of SZ [65]. Therefore, SZ is the second most common misdiagnosis of BD patients [59]. Additionally, the absence of (hypo-)manic symptoms when consulting a physician seems to be another cause of misdiagnosing BD with either MDD or SZ. According to Watanabe et al., approximately half of BD patients with an early correct diagnosis experienced mixed state or (hypo-)manic symptoms at the first visit whereas this was the case for only 5% of BD patients without an early correct diagnosis [56].

[4] In this study, an early correct diagnosis was defined as receiving a correct diagnosis within one year after the first visit to a medical institution.

1.2.5 Consequences of Misdiagnosis

A misdiagnosis is associated with a number of consequences for patients themselves as well as their caregivers and society. For patients, the greatest concern is the lack of appropriate treatment. As each psychiatric disorder requires its own specific therapeutic intervention, a misdiagnosis implies that the physician will decide on the best therapeutic intervention following guidelines for this misdiagnosis instead of the correct diagnosis. Consequently, it is likely that the patient's symptoms will not be alleviated by this therapeutic intervention, leading to a decreased quality of life among misdiagnosed BD patients compared to those who received a correct diagnosis [66-69].

Moreover, misdiagnosis could further deteriorate the condition of the patient as some therapeutic interventions include the administration of drugs with, on the one hand, an augmenting effect on symptoms of another disorder or, on the other hand, major side-effects. For instance, antidepressant monotherapy, the golden standard to treat MDD patients, is suggested to worsen the condition of BD patients by accelerating mood cycling and triggering instability [70-72], resulting in an increase of hospitalisation rates and suicide attempts [73]. Keeping in mind that one third of all suicidal acts in psychiatric patients occurs within the first year of illness, the importance of a fast and accurate diagnosis and thus treatment is highlighted by the significant reduction of the incidence of suicide in BD patients receiving appropriate treatment [74, 75].

Additionally, McCombs et al. demonstrated that, after one year, patients with recognised BD are on average US$1717 less costly than patients with unrecognised BD [76]. All studies examining the cost of unrecognised BD conclude that fast and accurate diagnosis of BD decreases the economic burden by lowering both direct (e.g. due to hospitalisation) and indirect costs (e.g. due to work loss) (for review see [77]). Furthermore, 30% of MDD and SZ patients do not respond to treatment strategies advised by guidelines for these disorders [78, 79]. However, evidence implies that approximately 40% of treatment-resistant MDD patients are in fact unrecognised BD patients, implicating that misdiagnosis of psychiatric disorders are partially causing concerningly low therapeutic efficacies of guideline treatments [80, 81]. Thus, accurate diagnosis could increase therapeutic efficacies of guideline treatments drastically.

As mentioned above, misdiagnosis of psychiatric disorders is mainly a result of the subjectivity hampering the diagnostic process. These high rates of misdiagnosis and more importantly their sequelae illustrate the dire need for objective diagnostic modalities allowing accurate and timely diagnosis of major psychiatric disorders, which could improve psychiatric healthcare by decreasing misdiagnosis rates, increasing therapeutic efficacies and paving the way towards personalised medicine.

1.3 Biomarker Discovery in Psychiatry

As the dire need for objective diagnostic modalities was and still is recognised by the scientific community, biomarker discovery in psychiatry was a subject undergoing intense study in the last decades. Many techniques, methods and sample sources have been investigated thoroughly for their potential to serve as a biomarker.

Neuroimaging has been extensively investigated because this technique allows to assess pathology-related alterations in a non-invasive manner. Although such alterations have been identified in psychiatric patients, both on a structural and functional level, many studies report conflicting findings (for review see [82, 83]). Moreover, many overlapping alterations have been identified in MDD, BD and SZ patients and would thus not allow differential diagnosis [84-86].

Another research area that has been extensively investigated in terms of biomarker discovery for psychiatric disorders is genomics. Family, twin and adoption studies revealed that the heritability of major psychiatric disorders is high (for review see [87-89]). This is specifically true for BD and SZ, in which the heritability rises to even up to 87% and 85%, respectively [90, 91]. Technological developments created the possibility to perform genome-wide association studies (GWAS) on large cohorts, which led to the identification of many risk genes (for review see [92]). Similar to neuroimaging, these genes are often implicated in multiple psychiatric disorders and are thus not fit for differential diagnosis.

1.4 Proteomics for Biomarker Discovery in Psychiatry

In the past years, proteomics gained interest in the field of biomarker discovery for psychiatric disorders since other approaches yielded disappointing results [93]. Proteomics, and more specifically non-targeted proteomics, offers a great platform for biomarker discovery by aiming to map the entire set of proteins in a desired cell, tissue or organism at a certain time point in a defined condition [94]. Therefore, profiling of the proteome might reveal underlying dynamic pathophysiological processes that may cause the varying symptomatology of psychiatric patients [95]. Furthermore, biomarker profiles can also increase our understanding of the molecular aetiology of psychiatric disorders and identify other disorders with similar molecular signatures, which in turn could even establish the basis to develop new therapeutics [96]. Hence, an increasing amount of proteomics studies was performed in the last decades, examining the proteome of brain tissue as well as peripheral fluids and cells (for review see [93, 97, 98]).

1.4.1 Proteomics of Peripheral Fluids and Cells

Although molecular fingerprinting of brain tissue could help to unravel underlying disease mechanisms, findings of such studies cannot be directly implemented as biomarkers since the collection of brain tissue is an extremely invasive procedure in living patients [99, 100]. Therefore, proteomics studies of fluids and cells, which are more feasible to collect in routine clinical practice, have been performed to investigate their potential as a sample source for biomarkers. Many studies focused on protein profiling of blood, mostly serum and plasma separately, because of the ease to collect the sample (for review see [93, 98, 101, 102]). Moreover, blood interacts with all tissues in the body and therefore contains important information regarding the health of our body, as demonstrated by the altered number of lymphocytes when our body is infected [103-105]. To a lesser extent, other fluids, such as cerebrospinal fluid (CSF), saliva, urine and sweat, were also investigated as potential sample sources for biomarkers (for review see [93, 95, 102, 106]). Although many studies made their data publicly available, comparison or a meta-analysis of the available data is difficult because of the large heterogeneity in experimental designs as well as statistical analysis [98]. The following sections provide a short overview of the best performing biomarker(s) (panels) that have been identified in previous studies.

1.4.1.1 Proteomics in Schizophrenia

The first blood-based test to aid in the diagnosis of SZ was established in 2010 by Schwarz et al. using the HumanMAP platform, a multiplexed immunoassay including 181 proteins at that time. The test consisted of a set of 34 proteins that were differentially expressed in serum of SZ patients compared to healthy controls. Although an algorithm using this biomarker panel could separate SZ patients correctly from healthy controls in 85% of the cases included in the cohort that was initially used for the discovery phase, this measure decreased significantly and was highly variable (40%-85%) when the algorithm was used to classify SZ patients and healthy controls in independent cohorts. Nevertheless, SZ patients could be discriminated from BD patients with a sensitivity and specificity of 86% and 78%

and from MDD patients with a with a sensitivity and specificity of 87% and 94%, respectively [107]. The sensitivity and specificity were not assessed in subsequent cohorts of the same psychiatric disorders, thus the variability of these measures when distinguishing SZ from other psychiatric disorders remains unknown.

Using the same platform, including 225 analytes at that time, Chan and colleagues identified 29 analytes that were differentially expressed in serum of SZ patients compared to healthy controls. 26 of those analytes were ultimately comprised in the biomarker panel. In the discovery cohort, this panel was able to discriminate SZ patients from healthy controls with a sensitivity and specificity of 90%. The biomarker panel's performance remained stable when it was used in a validation cohort [108].

1.4.1.2 Proteomics in Major Depressive Disorder

Papakostas and colleagues established an algorithm that is able to distinguish MDD patients from healthy controls with a sensitivity of 91,7% and a specificity of 81,3%, which remained stable in larger and independent cohorts [109]. This algorithm was upgraded in 2015 by including gender and body mass index (BMI) as interaction terms and normalising cortisol levels, which increased the sensitivity and specificity to 94% and 92%, respectively. The algorithm computes a MDDScore, which ranges from one in healthy controls to nine in depressed patients, based on the expression levels of nine serum analytes. Moreover, the MDDScore proved to be independent of antidepressant treatment, which allows it to be used in any patient [110]. However, the performance of the algorithm when distinguishing MDD patients from patients with another psychiatric disorder was not assessed, which is the ultimate goal in clinical practice.

Using the same multiplexed immunoassay as Schwarz and Chan et al. [107, 108], Stelzhammer and colleagues identified 11 serum analytes that could separate MDD patients from healthy controls with a sensitivity of 89% and a specificity of 95%. As Stelzhammer and colleagues assessed differences between MDD patients and healthy controls in two cohorts, differentially expressed analytes were only withheld when the differences were reproducible in both cohorts. Moreover, they concluded that two of these analytes were positively associated with symptom severity in cohort 1, but not in cohort 2 [111].

1.4.1.3 Proteomics in Bipolar Disorder

Biomarker panels to discriminate BD patients from healthy controls and other psychiatric patients are very scarce. In 2016, Haenisch and colleagues identified 20 differentially expressed analytes in a discovery phase, validated them in a validation phase and tested their potential for clinical usefulness in an application phase. In the discovery phase, 87 analytes were detected in serum of BD patients and healthy controls using the Human DiscoveryMap platform. The patients and healthy controls were derived from eight cohorts recruited at five different sites. The differentially expressed analytes could distinguish BD patients from healthy controls with an area under the curve (AUC) of 0,90 when they were combined in a biomarker panel. The performance of the biomarker panel remained stable when it was tested in an independent cohort despite the fact that data were only available for 16 of the 20 analytes selected in the discovery phase (AUC = 0,92). Furthermore, the predictive performance of the biomarker panel was assessed in a cohort containing 102 MDD patients of which 12 experienced a (hypo-)manic episode within two years after blood collection and were subsequently diagnosed with BD (AUC = 0,84). This suggests that the biomarker panel is clinically useful even though the sample size is too small to make solid conclusions [112].

1.4.1.4 Proteomics Based on the Comparison of Differential Diagnoses

Various studies focused on biomarker discovery for multiple psychiatric disorders with similar clinical presentations at the same time. However, such studies selected biomarker candidates only based on their discriminatory potential between psychiatric patients and healthy controls [107, 113, 114]. To the best of my knowledge, only three studies compared protein expression levels of patients with differential psychiatric disorders directly with each other. More specifically, they all investigated differential expressions across the proteome of MDD and BD patients [115-117]. The research papers reporting the results of these studies were all published in the last five years, indicating that this comparative approach only emerged very recently.

In 2015, Chen and colleagues identified 25 proteins that were differentially expressed in plasma of MDD patients compared to BD patients by using a non-targeted proteomics approach. Validation of three of those proteins using Enzyme-Linked Immuno Sorbent Assay (ELISA) yielded quantitative results that were consistent with those of the non-targeted approach [115]. Another study using a non-targeted proteomics approach was conducted in 2017 and identified 1012 proteins in plasma of the same patient populations and healthy controls. In total, 116 differentially expressed proteins were identified across the three comparisons (MDD/control, BD/control, MDD/BD), of which only 9 were differentially expressed between MDD and BD patients [116]. There was no overlap between these 9 proteins and the 25 proteins that were identified by Chen and colleagues [115, 116]. A very recent study, published in April 2020, profiled expression levels of proteins in serum of MDD and BD patients. Of the 268 proteins that were considered for analysis, 14 showed to be differentially expressed. After correction for multiple testing, only three of those remained significantly differentially expressed [117]. Interestingly, one of the three proteins that remained differentially expressed after correction for multiple testing was also identified as differentially expressed in plasma of these patient populations by Chen et al. [115, 117]. None of these studies assessed the performance of the identified biomarkers when distinguishing patients with differential diagnoses from each other.

1.4.2 Limitations of Previous Proteomics Studies

Clearly, many studies focused on biomarker discovery for major psychiatric disorders using various techniques, methods and sample sources. However, none could identify a biomarker panel that is useful in clinical practice. Promising biomarkers identified in a research environment showed to be non-specific in a clinical setting and thus were of no use for clinical diagnosis of major psychiatric disorders. This specifically appeared to be true for single biomarkers, possibly because of their inability to represent complex underlying disease mechanisms [118, 119]. Therefore, psychiatry is one of the few fields lacking biological measures for diagnosis.

In my opinion, the greatest limitation of many proteomic studies focusing on biomarker discovery for psychiatric disorders is the fact that they identify biomarker(s) (panels) based on their ability to distinguish patients suffering from a certain disorder and healthy controls. However, the challenging aspect of diagnosing psychiatric disorders is distinguishing differential diagnoses from each other, rather than deciding whether a patient is healthy or has a certain disorder. This might be causing that promising biomarker(s) (panels) are not specific and therefore not useful in clinical practice. Moreover, most studies do not take into account the different states associated with a certain psychiatric disorder. As these states present with a diverse symptomatology, underlying molecular processes causing these symptoms might be distinct. Therefore, proteins related to those processes might fluctuate resulting in state-specific biomarkers [119, 120]. For instance, BD is characterised by cycling between manic and depressive states, and SZ patients do not always present with psychotic symptoms. Furthermore, the majority of proteomic studies used a targeted approach, which may be biased by the included proteins and does not allow to quantify expression levels of most proteins. Illustratively, the

HumanMAP platform, which is widely used for biomarker discovery in psychiatry, is mainly focused on the quantification of immune mediators and cytokines [113]. Consequently, such approaches impede the disclosure of dysregulated pathways and protein networks outside the scope of the included proteins. Moreover, studies implementing a non-targeted proteomics approaches all used a much smaller sample size compared to when targeted approaches were used, resulting in a decreased power. The same holds true for studies comparing differential diagnoses directly with each other instead of healthy controls.

Additionally, many of these studies focused on protein profiling of serum or plasma. However, a recent study highlighted that the use of PBMCs might be preferred over plasma for biomarker discovery because i) "cellular proteomes in general deliver a higher number of proteins than plasma" (see Figure 3), ii) "the focus of detected proteins in cellular proteomes is substantially different and significantly richer as a source of biomarkers" and iii) low-abundant proteins, which are more likely to be relevant biomarkers, are masked by high-abundant proteins in plasma. Furthermore, the proteome of PBMCs and plasma is significantly different (see Figure 3) [121]. Hence, biomarker discovery might be more successful using PBMCs instead of plasma.

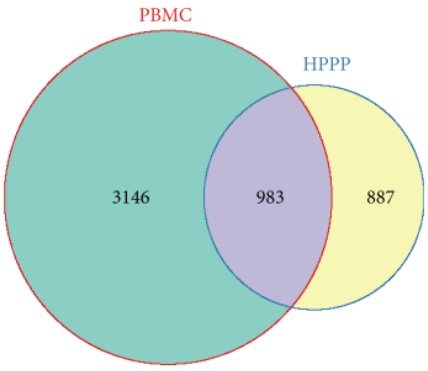

Figure 3 Number of proteins identified in PBMCs and plasma. The Venn diagram displays the number of proteins identified in PBMCs and plasma (HPPP) and the number of proteins overlapping between the proteome of PBMCs and plasma. Abbreviations: HPPP, Human Plasma Proteome Project; PBMCs, peripheral blood mononuclear cells [121]

1.5 PBMCs for Biomarker Discovery in Psychiatry

PBMCs comprise all blood cells containing a round nucleus, including lymphocytes, monocytes and dendritic cells, and are therefore essential players of the immune system [122]. In terms of technical and practical matters, PBMCs are an excellent source for biomarkers because they are abundantly present in a venous blood sample. Per millilitre whole blood, PBMC counts range from 0,5 to 3 million [123-126]. Moreover, a venous blood sample is collected via a minimally invasive technique which is routinely used in clinical practice. Therefore, a diagnostic test based on PBMCs could be implemented in daily routine very fast once a biomarker panel is developed and validated. Furthermore, proteomics analysis of PBMCs yields a high number of proteins (Figure 3), increasing the chances to identify biomarkers compared to plasma or serum [121]. Other than the beneficial characteristics of PBMCs in terms of technical and practical matters, they are specifically interesting for biomarker discovery of brain disorders as they are proposed to be a neural probe [127]. This way, identified biomarkers might be reflecting pathophysiological processes occurring in the brain, resulting in an increased understanding of disease mechanisms [128, 129].

1.5.1 Bidirectional Communication between PBMCs and the Brain

PBMCs manifest similar gene and protein expression profiles as the brain. For example, a study of Rollins et al., comparing gene expression levels in PBMCs and post mortem brain tissue of the same individual, revealed that a number of genes showed similar expression levels in PBMCs and brain tissue [100]. Furthermore, some of these similarly expressed genes and proteins, such as receptors for brain derived neurotrophic factor, serotonin and dopamine, have already been suggested to be involved in psychiatric disorders [127]. Similarities in gene expression profiles in PBMCs and brain tissue may be a result of the bidirectional communication occurring between them.

On the one hand, the brain influences the immune system, including PBMCs, through various mediators, such as neurotransmitters and hormones [130, 131]. For instance, gamma-aminobutyric

acid (GABA) and glucocorticoids regulate cytokine production in PBMCs by binding to their receptors on the cell membrane of PBMCs and activating signalling cascades. This way, GABA and glucocorticoids can change gene expression and modulate the immune system [130, 132]. Additionally, the brain also affects PBMCs similarly via cytokines, which are important signalling molecules of the immune system [133]. Although they are mainly secreted by immune cells, including PBMCs, cytokines are also secreted by microglia, astrocytes and neurons [134, 135].

On the other hand, PBMCs influence the brain via various signalling molecules of the immune system, for which microglia, astrocytes and neurons express receptors on their cell membrane [136, 137]. Activation of these receptors allows signalling molecules to have an impact on normal functions of the central nervous system, such as the regulation of sleep [138] and neuronal development [139, 140]. These signalling molecules communicate with the central nervous system via neural routes, passive and active transport across the blood brain barrier and interaction with cells of the blood brain barrier [141].

1.5.2 Involvement in Psychiatric Disorders

PBMCs are specifically of great interest for biomarker discovery in psychiatry as decades-old observations implicate altered PBMC features in psychiatric patients. These alterations entail both macroscopic characteristics, such as deviant amounts of PBMCs [142], as well as molecular processes, such as an impaired oxidative metabolism [143]. As new technologies were developed, the knowledge about PBMC alterations in psychiatric patients expanded by more recent studies focusing on all sorts of omics [144-146].

Coppens and colleagues investigated whether or not PBMCs could serve as a source of biomarkers for differential diagnosis of psychiatric disorders with similar clinical presentations, namely MDD patients and BD patients in a depressive state (BD-D), on the one hand, and BD patients in a manic state (BD-M) and SZ patients, on the other. The study identified proteins that are differentially expressed in these patients as well as compared to healthy controls, which implicates that the identified proteins are also pathology related [147]. Hence, the results of this study suggest that PBMCs offer a great source of biomarkers for psychiatric disorders. Furthermore, other studies reported alterations that could be related to hypotheses of underlying disease mechanisms [148-152]. Illustratively, Torres and colleagues noted that expression levels of receptors involved in dopaminergic pathways are aberrant in PBMCs of BD and SZ patients [148, 149].

Even more interesting is that some PBMC alterations have been shown to be condition-specific. For instance, Gurvich et al. observed episode-specific gene expression shifts in PBMCs of BD patients [153]. Another recent study reported that altered DNA methylation, and thus also expression, of stress related genes in PBMCs of MDD patients may only be present in patients with serious suicide ideation and not in other MDD patients [154]. Consequently, condition-specific alterations of PBMC features may result in an increased efficiency to detect conditions in need of extra care or treatment.

1.6 Aims of the Project

Objective diagnostic biomarkers are currently lacking in the field of psychiatry. Therefore, this thesis aimed to identify biomarker candidates for differential diagnosis of major psychiatric disorders via protein profiling. More specifically, differential diagnosis of MDD and BD-D patients and differential diagnosis of BD-M and SZ patients were prioritised because of the high misdiagnosis rates associated with these afflictions.

A first step towards this aim was to identify which sample source offers the best characteristics for biomarker discovery. Samples of healthy volunteers were used to compare the proteome of plasma, the golden standard in biomarker discovery, and PBMCs, proposed to be neural probes, in terms of quantitative and qualitative endpoints. First, the protocol for sample preparation of plasma was optimised because an optimised protocol for this purpose was not yet available in the host laboratory. However, this optimisation phase did not provide a solid optimised protocol for sample preparation of plasma as the goal of this phase was to optimally utilise the available resources rather than to create an optimised protocol that is as strong as possible in technical terms.

Additionally, data of a proof-of-concept study using protein profiling of PBMCs of the same experimental groups as described earlier were analysed. Biomarker candidates were identified and subjected to pathway and network analysis to obtain a better understanding of affected pathways, cellular functions or systems. As this research was conducted in a highly exploratory approach, data were obtained in a hypothesis-free manner.

2. Methodology

2.1 Initial Experimental Set-up

2.1.1 Optimisation Protocol for Sample Preparation and Quantitative Proteomics

2.1.1.1 Sample Collection

Technical samples, obtained from healthy volunteers, were collected between 7:30 and 10 AM in 10 mL K_2-EDTA-coated collection tubes (REF 367525, BD Vacutainer®, New Jersey, USA). 5 mL whole blood was pipetted onto Histopaque-1077 (Sigma-aldrich, Missouri, USA) to collect PBMCs. The buffy coat, containing approximately one million PBMCs per millilitre blood [123-126], was collected following gradient centrifugation for 20 minutes at 700g and 20°C without brakes. Next, phosphate buffered saline (PBS; Thermo Fisher Scientific, Perth, United Kingdom) was used to wash the buffy coat twice. After discarding the supernatant, dry cells were stored at -80°C until further analysis (Figure S1 in Supplementary materials). The remaining 5 mL whole blood was centrifuged for 10 minutes at 3.6 rpm and 4°C to obtain plasma, which was aliquoted and stored at -80°C until further analysis (Figure S2 in Supplementary materials).

2.1.1.2 Sample Preparation of Plasma

For this study, three different aliquots of the same test subject were prepped. Of those, one aliquot was prepped in triplicate (Figure 4). As an optimised protocol for sample preparation of plasma was not yet available, a general sample preparation protocol for plasma was optimised by assessing the quality of some commercially available products [155, 156].

Although non-targeted proteomics offers a great platform for biomarker discovery, the large dynamic range of protein abundances in plasma impedes mass spectrometry (MS)-based protein profiling [157, 158]. Therefore, depletion of highly abundant proteins is necessary to allow detection of proteins with lower abundances [159, 160]. Moreover, Geyer and colleagues stated that less abundant proteins are possibly more likely to be identified as biomarkers than highly abundant proteins [161]. To achieve depletion of highly abundant proteins, the easiest and most commonly used method involves the use of solid phase extraction columns. These columns can be designed specifically for the depletion of highly abundant proteins in plasma, of which many are commercially available [160, 162]. Two of such commercially available columns are the ProteoSpin™ Abundant Serum Protein Depletion Kit (also fit for plasma, Norgen, Ontario, Canada) and the Pierce™ Top 12 Abundant Protein Depletion Spin Column (Thermo Scientific, Massachusetts, USA). Although the latter is the most well-known and widely used column for this purpose, it is three times more expensive than the former. Therefore, the performances of both columns were compared to assess whether or not the Pierce™ column is worth its price[5]. Samples were depleted following manufacturers' instructions.

As this study involves the quantification of proteins, an equal amount of proteins for each sample should be analysed. Hence, the determination of the protein concentration by an RC DC protein assay (Bio-Rad, California, USA) allowed us to calculate the necessary volume of each sample. Subsequently, the samples were reduced and alkylated to allow separation by the LC column. Finally, peptides were generated by digesting the proteins present in the sample via an appropriate enzyme or chemical. To

[5] The Pierce™ column would only be considered for use in the biomarker discovery phase when it significantly outperforms the ProteoSpin™ column. In case both columns perform equally well, the ProteoSpin™ column will be selected for the biomarker discovery phase based on the cost of the column.

perform these last two steps, the ProteoSpin™ On-Column Proteolytic Digestion Kit (Norgen, Ontario, Canada) was used according to the manufacturer's instructions. Trypsin was used to digest proteins because of its high specificity and ease to use [163]. All prepped samples were stored at -20°C before LCMS analysis to mimic the experimental situation of the biomarker discovery phase as much as possible.

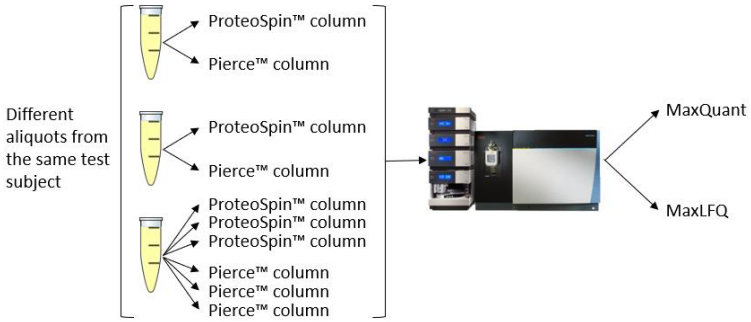

Figure 4 Visual representation of the experimental design for comparing depletion columns.
Abbreviations: PBMCs, peripheral blood mononuclear cells

2.1.1.3 Sample Preparation of PBMCs

Three aliquots, two of which were derived from the same test subject, were prepped using the standard protocol for sample preparation of cells (Figure 5). PBMCs of technical samples were solubilised in a protein extraction buffer (RIPA buffer (150 mM NaCl, 50 mM Tris, 0,5% sodium deoxycholate, 1% NP-40) containing 1% SDS) using a sonicator (UP50H, Hielscher Ultrasonics, Germany). To ensure equal amounts of proteins in every sample, the necessary volume of each sample was calculated based on the results of a RC DC protein assay (Bio-Rad, California, USA) before proceeding to the next step. Subsequently, the ProteoSpin™ On-Column Proteolytic Digestion Kit (Norgen, Ontario, Canada) was used for reduction, alkylation and digestion of proteins in the samples following the manufacturer's instructions. All prepped samples were stored at -20°C before LCMS analysis to mimic the experimental situation of the biomarker discovery phase as much as possible.

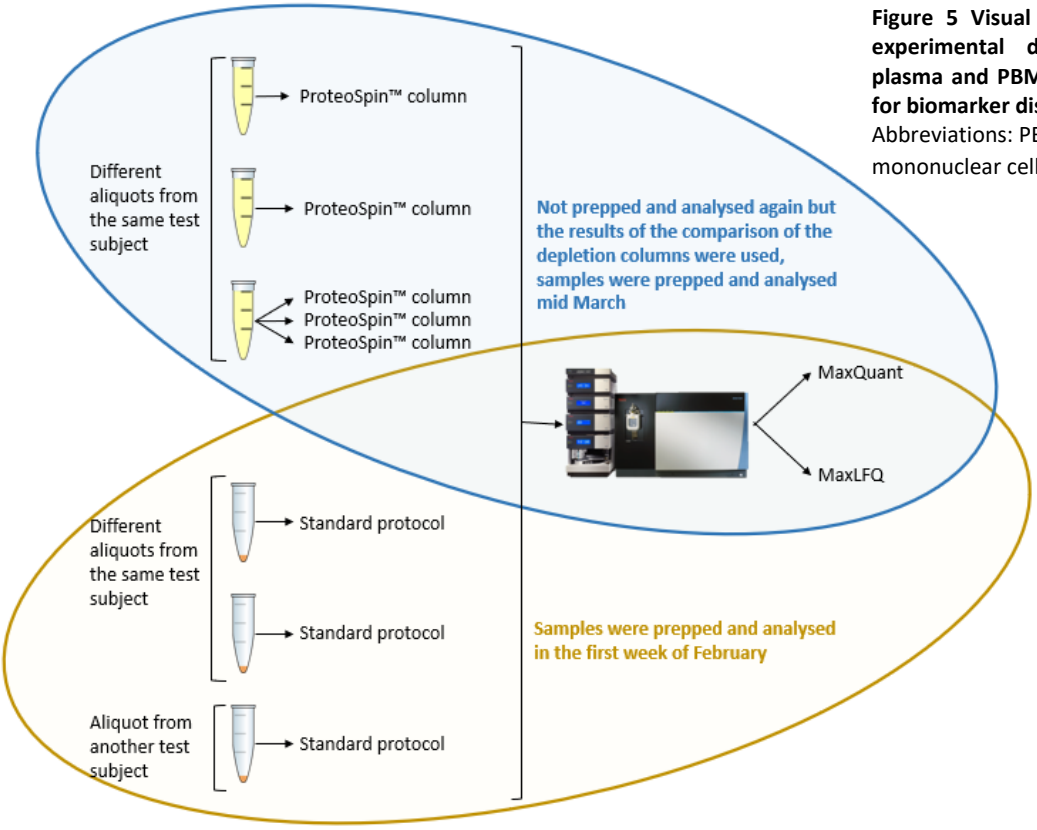

Figure 5 Visual representation of the experimental design for comparing plasma and PBMCs as a sample source for biomarker discovery.
Abbreviations: PBMCs, peripheral blood mononuclear cells

2.1.1.4 LCMS

All samples were loaded *in triplo* onto the LC fractionation system (Dionex ULTIMATE 3000, Thermo Scientific, Massachusetts, USA), which consists of a nano-LC C18 column (200 Å, 2 µm, 75 µm × 25 cm), and eluted with a LC gradient of 155 minutes. As this column is coupled online to a QExactive™-Plus Orbitrap MS (Thermo Scientific, Massachusetts, USA), the eluents were automatically infused to the MS with a capillary at 1.7 KV on a nano-ESI source at a flow rate of 300 nl/min.

Additionally, the length of the LC gradient that is used to elute peptides from the column was increased to try to increase the number of proteins that could be retrieved from the samples as this parameter already proved to have an effect on this number [164, 165]. The residuals of the two plasma samples with the highest number of identifications that were depleted with ProteoSpin™ were used to assess this effect. For this purpose, the above described LCMS procedure was repeated with a varying length of the LC gradient (180 minutes, 240 minutes, 300 minutes, 360 minutes and 440 minutes).

2.1.1.5 Protein Identification and Quantification

Data were acquired for a selected mass range of 350-1800 m/z at the MS1 level with a resolution of 140,000 and at the MS2 level with a resolution of 17,500 and were analysed by MaxQuant Software (Open Source, [166]). Andromeda was used as search engine to generate peptide sequences based on the human UniProt/SwissProt database and to identify tandem mass spectra with a confidence over 99%. Settings of MaxQuant software are listed in Table S1 in the Supplementary materials.

Moreover, the effect of analysing the data with the MaxLFQ algorithm on the variation was examined by analysing all samples with this algorithm and the standard MaxQuant algorithm (see Figures 4 and 5). The MaxLFQ algorithm was specifically designed for label-free quantification MS and showed to quantify protein abundances more accurately than other methods [167, 168]. MaxLFQ implements a delayed normalisation via a global optimisation procedure after the summation of peptide intensities. Based on the assumption that most protein expression levels do not change between two conditions, normalisation factors are determined by aiming for the least overall proteome variation. Proteins are quantified by calculating protein ratios between any pair-wise comparison of samples that were analysed. Each pair-wise protein ratio is defined by the median of the peptide ratios derived from the same pair-wise comparison. This way, protein quantification is based on the maximum possible information that can be retrieved from the data. A more detailed explanation of MaxLFQ is available in the original manuscript [168].

2.1.1.6 Data Analysis

Graphad Prism was used to perform statistical analysis with a statistical significance set at $p < 0,05$. The Shapiro-Wilk test was used to check for normality of data because it proved to be the most sensitive normality test. However, a study assessing the sensitivity of this test by subjecting it to non-normal data concluded that it failed to reject the null hypothesis[6] when the sample size was below 40 [169]. Therefore, statistics were performed using a conservative approach and datasets with a sample size below 40 were analysed using non-parametric tests. Moreover, this conservative approach is in line with the fact that this thesis was focused on detecting large differences as conservative approaches tend to be more careful with respect to detecting significant differences [170]. The depletion columns were compared to determine whether or not the Pierce™ column is worth its price, which Is three times the price of the ProteoSpin™ column and thus would only be considered when

[6] The null hypothesis of the Shapiro-Wilk test states that data are normally distributed.

there are large differences. Significant differences were determined using the Wilcoxon signed-rank test in case the depletion columns were compared and the Mann Whitney U test in case plasma was compared to PBMCs. Data are presented as median [lower limit – upper limit].

To validate the method with the highest qualitative return, qualitative and quantitative endpoints were compared. Data was not transformed to capture real variation. The coefficient of variation (CV), defined as the ratio of the standard deviation to the mean, was calculated at multiple levels using Microsoft Excel. The variation on the intensity of each protein was determined across separate sample preparations of the same aliquot (inter-prep variability), across different LCMS runs (inter-run variability) and across different test subjects (inter-individual variability) (see Figure 6). Moreover, the range of the inter-run variability was determined for each separate protein by subtracting the maximum with the minimum inter-run CV. The number of proteins with a CV below 20% was also calculated at all levels. The cut-off was set at 20% as this is frequently used for diagnostic assays [171]. Additionally, the total number of distinct proteins identified per sample was calculated by summing up all individual proteins that were identified in at least one run. To assess the detection stability, the number of proteins that were identified in each run of a sample was also calculated. This is an important measure because detection in multiple runs is necessary to reach quantitative precision. These parameters were all assessed when using the standard MaxQuant algorithm as well as the MaxLFQ algorithm.

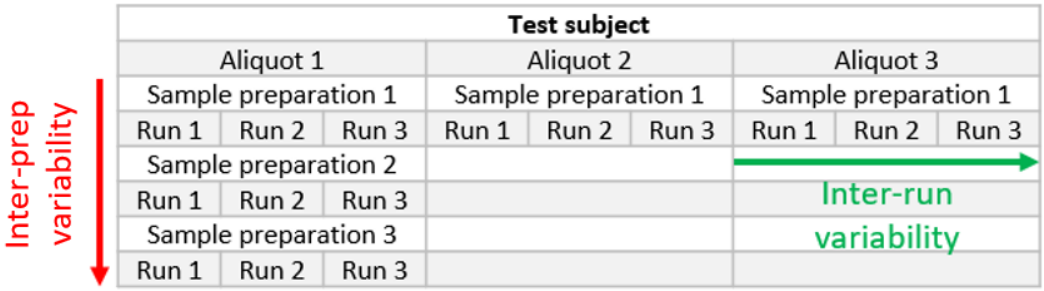

Figure 6 Visual representation of the calculation of the variation at multiple levels.
Abbreviations: prep, preparation

2.2 Replacement Assignment

Unfortunately, the initial experimental set-up could not be completed due to the COVID-19 outbreak. Therefore, this thesis included analysing a dataset containing protein expression levels of PBMCs of the same populations as the initial experimental set-up. These data were obtained in a proof-of-concept study of which the experimental set-up is copied from the original manuscript with approval of the authors [147]. The original manuscript is supplied in the supplementary materials.

2.2.1 Patient Selection and Sample Collection

MDD patients (n = 5), BD-D patients (n = 3), BD-M patients (n = 4) and SZ patients with active positive symptoms (n = 4) were recruited from UPC Duffel. HCs were recruited via advertising and matched to patients based on age and gender. The study was approved by the Ethics committee of Antwerp University Hospital and UPC Duffel. PBMCs were collected as described in section 2.1.1.1.

2.2.1.1 Inclusion Criteria

Subjects must meet the following criteria before entering the study:

1. Be a man or a woman between 18 and 55 years, inclusive

2. Have signed an informed consent document (indicating that they understand the purpose of, and procedures required for the study and are willing to participate in this study.
3. Be medically stable on the basis of physical examination and vital signs performed at Screening.
4. Be medically stable on the basis of clinical laboratory tests performed at Screening. If the results of the serum chemistry, haematology, or urinalysis are outside the normal reference ranges, the subject may be included only if the investigator judges the abnormalities or deviations from normal to be not clinically significant or to be appropriate and reasonable for the population under study.
5. Be willing and able to adhere to the prohibitions and restrictions specified in the protocol.
6. For patients: Be diagnosed with either MDD, BD (type 1 or 2) or SZ according to the diagnostical MINI-interview
7. For inclusion in one of the four patient groups, patients additionally must have:
    - Group 1 major depressive disorder patients: a Hamilton Depression Rating Scale (HDRS) score of 14 or higher
    - Group 2: Bipolar disorder patients, depressed phase a HDRS score of 14 or higher
    - Group 3: Bipolar disorder patients, (hypo)manic phase: a Young Mania Rating Scale (YMRS) score of 13 or higher
    - Group 4: schizophrenia patients: A total score of ≥14 on the positive scale of the Positive And Negative Syndrome Scale (PANSS) and at least a score of 5 on 1 item or a score of 4 on 2 "psychotic" PANSS items P2, P3, P5 or G9 at Screening.
8. Individuals have to be physically healthy on the basis of clinical judgment of the investigator.

2.2.1.2 Exclusion Criteria

Potential subjects who meet any of the following criteria will be excluded from participating in the study:

1. Applicable to the control group:
    a. Personal history of psychotic disorder or mood disorder
    b. Family history of psychotic or mood disorder in first-degree relatives
2. Has a history of drug or alcohol dependence according to DSM-5 criteria, except nicotine or caffeine, within 6 months before screening.
3. Has history of (co-morbid) somatisation or mood disorder according to DSM-5 criteria within 6 months before screening.
4. Has a positive test result for drugs of abuse or for alcohol at screening or test day.
5. Female subjects only: is pregnant or breastfeeding
6. Has a history of chronic or acute physical illness associated with abnormal immune changes within the 2 weeks before the study.
7. Leucocytosis (i.e., white blood cell count ≤ 11 x109 /L) on screening and test days.
8. Serology positive for hepatitis B surface antigen (HBsAg), hepatitis C antibodies, or HIV antibodies at screening.
9. Has a medical history of any auto-immune disorder or chronic inflammatory disease.
10. Has received electroconvulsive therapy in the last 6 months.
11. Is currently enrolled in a study with an investigational study drug.
12. Has any condition that, in the opinion of the investigator, would compromise the wellbeing of the subject or the study or prevent the subject from meeting or performing study requirements.

2.2.2 Quantitative Proteomics

Multiplexed iTRAQ (isobaric mass-tag labelling for relative and absolute quantitation) LCMS was performed as described hereafter. PBMC samples were solubilised in a protein extraction buffer (composition: 8 M urea, 2 M thiourea, 0.1% SDS and 50 mM triethylammonium bicarbonate). Next, protein concentrations were quantified using RC DC protein assays (Bio-Rad; California, USA). Equal amounts of proteins from each sample were then reduced by tris-2-carboxyethyl phosphine and alkylated by 5-methyl-methanoethiosulphate and finally subjected to trypsin digestion. The resulting peptides from each sample were labelled using iTRAQ reagents (Sciex, Massachusetts, USA) following the manufacturer's instructions.

PBMC samples of HC and patient samples were spread randomly across three different octaplex iTRAQ LC runs. To improve LC-MS/MS proteome coverage, samples were subjected to a 2D-LC fractionation system (Dionex ULTIMATE 3000, Thermo Scientific, Massachusetts, USA). Peptide mixes were fractionated on a strong cationic exchange chromatography column (1 mm x 150 mm polysulfoethyl Aspartamide (California, USA, Dionex)) separated subsequently carried on a nano-LC C18 column (200 Å, 2 µm, 75 µm × 25 cm). The nano-LC is coupled online to a QExactive™-Plus Orbitrap (Thermo Scientific) mass spectrometer (MS). The nano-LC eluents were infused to the Orbitrap mass-spectrometer with a capillary at 1.7 KV on a nano-ESI source at a flow rate of 300 nl/min.

Data dependent acquisition in positive ion mode was performed for a selected mass range of 350-1800 m/z at the MS1 level with a resolution of 140,000 and at the MS2 level with a resolution of 17,500. The raw data were analysed by Proteome Discoverer 2.1 Software (Thermo Scientific) using Sequest HT as the search engine against the human UniProt/SwissProt database. The threshold of confidence was set above 99% ensuring a false discovery rate of less than 1%. The list of identified proteins, containing iTRAQ ratios of expression levels over control samples, was generated. Proteome Discoverer 2.1 employs a global analytical methodology.

2.2.3 Data Analysis

2.2.3.1 Demographics

For demographics, group mean differences were calculated by ANOVA with Tukey honest significant differences (HSD) post-hoc comparisons for numerical data and with Fisher exact test for categorical variables (gender, smoking status). All analyses were performed using JMP® version 13 (SAS, Cary, North Carolina 27513, USA). Demographics are presented as mean ± standard deviation (SD).

2.2.3.2 Identification of Biomarker Candidates for Differential Diagnosis

Because of the small sample size, protein expression levels were only analysed when the protein was detected in all HC and patient samples to limit false discoveries. Raw iTRAQ ratios (condition/standard) were accumulated into mean abundance ratios per experimental group. Subsequently, these mean abundance ratios were used to calculate the fold change expression ratios of each patient group over HCs. Furthermore, the fold change expression ratios were also calculated for MDD and BD-D patients, on the one hand, and BD-M and SZ patients, on the other hand. All fold change expression ratios were $\log_2$ transformed to centre them around zero and thus reduce skewness. Finally, two lists of biomarker candidates were assembled: one for differential diagnosis of MDD and BD-D patients and one for differential diagnosis of BD-M and SZ patients. These lists contain proteins of which the $\log_2$ fold change expression ratio is larger than two SD to i) the patient/patient mean $\log_2$ fold change expression ratio and ii) the patient/HC mean $\log_2$ fold change expression ratio in at least one of the two patient groups

(Figure S3 in Supplementary materials) [172]. This way, biomarker candidates are only considered when they are pathology-related. VennPlex was used to visualise this *modus operandi* [173].

2.2.3.3 Pathway Analysis

Protein expression values were further analysed via pathway analysis to identify enriched gene sets, the topology-free variants of pathways. As many pathway analysis methods are available, selecting the best methods for your study can be challenging and should be considered carefully. In this thesis, pathway analysis was performed by using the CAMERA method, or Correlation Adjusted MEan RAnk gene set test, because of the following reasons. First, CAMERA takes into account inter-gene correlations [174]. This is an important aspect because various studies have demonstrated that the false discovery rate can drastically inflate when these correlations are ignored, as is the case in many other pathway analysis methods [175-177]. Second, p-values of pathways generated by CAMERA are uniformly distributed between 0 and 1 [174], in contrast to the skewed null distribution of p-values generated by other methods for pathway analysis [178]. Illustratively, CAMERA identifies a number of gene sets to be differentially enriched that is close to the expected (5%) when samples are randomly labelled and significance is set at 0,05 [179]. Moreover, this number of gene sets predicted to be differentially enriched showed to be independent of the sample size whereas many other pathway analysis methods generated an increasing number of gene sets predicted to be differentially enriched when the sample size was enlarged [180]. Third, Maleki and colleagues demonstrated that the number of gene sets predicted to be differentially enriched by this method is reasonable while this number is much larger for other methods, of which many are known to generate false positives [178, 181].

Pathway analysis was performed using R Statistical Software, more specifically the *limma* package [182]. This package includes a function that automatically performs pathway analysis via the CAMERA method [174]. Gene sets were downloaded from MSigDB[7] because this database is a collection of gene sets derived from many other well-known databases, such as Reactome Pathways and KEGG, and therefore includes more information than a single database [183]. Using the *gmtPathways* function, which requires the *Rcpp* and *fgsea* packages, gene sets were loaded into an R object. Subsequently, this object was used to map genes, corresponding to the identified proteins, to the gene sets they belong to via the *id2indices* function. The *camera* function also requires an expression matrix, containing expression values of each protein in each sample, a design matrix, specifying to which experimental group a sample belongs, and a contrast matrix, defining the comparison of interest. Moreover, the inter-gene correlation was set at 0,01, as recommended in the R documentation. CAMERA assigns the direction of regulation (up- or down-regulated), a p-value and a q-value, the p-value corrected via the Benjamini–Hochberg algorithm, to each gene set included in the database that was used.

As the *camera* function does not take into account multiple contrasts simultaneously (e.g. patient/patient and patient/control), enriched gene sets were identified in the same manner as biomarker candidates were identified. Thus, a gene set should be enriched (q-value below 0,05) when i) expression values of patient groups were compared with each other and ii) expression values of HCs and patients were compared in at least one of the two patient groups. Interactions between significantly enriched gene sets were visualised via Cytoscape 3.7.2 [184] using the Cytoscape app Enrichment Map [185]. Clusters within these interaction networks were identified and annotated by

[7] Accessed on June 1ste 2020

AutoAnnotate and WordCloud [186, 187]. Overarching terms were manually added via Microsoft PowerPoint.

2.2.3.4 Network Analysis

Although pathway analysis is a great way of gaining more meaningful and interpretable information from a list of genes or proteins, it only considers canonical pathways or gene sets. As network analysis offers the opportunity to implement non-canonical information, the lists of biomarker candidates for differential diagnosis were also subjected to this type of analysis by using Cytoscape 3.7.2 [184]. Two networks, one based on biomarker candidates for differential diagnosis of MDD and BD-D patients and one based on biomarker candidates for differential diagnosis of BD-M and SZ patients, were created with information gathered via the STRING database. This database generates protein interactions based on experimental data as well as computational prediction efforts and therefore creates networks based on protein interactions at the widest scope [188]. The Cytoscape app stringApp is required to implement this information in Cytoscape [189].

Network characteristics were assessed with the Cytoscape tool NetworkAnalyzer [190] and cytoHubba, another Cytoscape app, was used to identify hubs and bottleneck proteins [191]. Nodes with a high degree (i.e. number of direct neighbours) are called hubs whereas bottleneck proteins are characterised by a high betweenness centrality (i.e. number of shortest going through that node). Both are considered to be essential elements of a network [192, 193]. However, there is no inherent threshold defining which node is a hub or a bottleneck [192]. Therefore, hubs and bottlenecks are often referred to as the top 5-20% proteins with the highest degree or betweenness centrality, respectively [193-197]. As this is a highly explorative study, the threshold was set at 20%. A node that reaches the threshold to be considered a hub as well as a bottleneck is referred to as a hub-bottleneck.

Additionally, cluster analysis of these networks was performed using the Cytoscape app Cluster ONE because this algorithm allows that one node, corresponding to one protein, can be part of multiple clusters, in contrast to other cluster algorithms where a node is assigned to just one of the clusters. This way, the Cluster ONE algorithm takes into account that one protein may have multiple functions and therefore clusters in a manner that is closer to the true biological situation [198]. Cluster analysis was performed with the default settings for all parameters, except for the edge weights. Edges were weighted using the STRING confidence score, which is automatically generated when the network is created and represents the likelihood for the interaction to be true. Finally, all clusters were subjected to functional enrichment analysis using the stringApp [189]. All proteins of the network were included as background. Terms were not filtered based on redundancy because this approach only withholds the most significant term(s) when there is a certain degree of overlap. However, the significance assigned to a gene set may be positively correlated with the size of the gene set [179]. Consequently, larger and more general gene sets are more likely to be withheld, leading to a loss of information about smaller and more specific gene sets.

3. Results

3.1 Choice of Depletion Column and Analytical Algorithm

3.1.1 Protein Yield

As a first optimisation strategy, two types of commercially available and widely used depletion columns were compared in terms of the number of distinct proteins that were retrieved. When a sample was run in triplicate, the median total number of distinct identifications was higher in samples depleted with the ProteoSpin™ column compared to in those depleted with the Pierce™ column (196 [191 – 198] vs 162 [158 – 169]; p = 0,0625; Figure 7a). Moreover, two analytical algorithms were compared using the same endpoints. The difference in the number of identifications between the two depletion columns attenuated when the samples were analysed using the MaxLFQ algorithm (139 [133 – 168] vs 132 [130 – 143]; p = 0,1875; Figure 7a). Furthermore, analysis with the MaxLFQ algorithm resulted in a borderline significant decrease in the total number of identifications in samples depleted with the ProteoSpin™ column (p = 0,0625; Figure 5a). This was also the case for samples depleted with the Pierce™ column (p = 0,0625; Figure 7a).

An important measure to consider when optimising a protocol for biomarker discovery is detection stability. Therefore, the number of distinct proteins that were detected in all runs per sample was determined. The number of proteins identified in all runs of a sample was borderline significantly higher in samples depleted with the ProteoSpin™ column compared to those depleted with the Pierce™ column (171 [165– 174] vs 139 [131 – 150]; p = 0,0625; Figure 7b). Furthermore, implementation of the MaxLFQ algorithm decreased the number of proteins identified in all runs of a sample for both depletion columns (116 [109 – 118], p = 0,0625 for samples depleted with the ProteoSpin™ column; 112 [105 – 121], p = 0,0625 for samples depleted with the Pierce™ column; Figure 7b). The difference in number of proteins identified in all runs between the depletion columns attenuated when the MaxLFQ algorithm was used (p = 0,3750; Figure 7b).

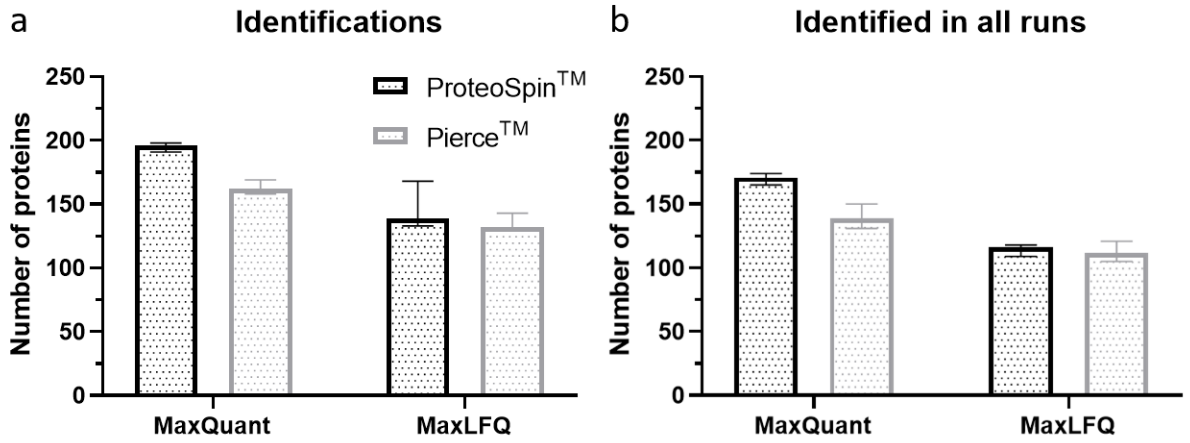

Figure 7 Number of proteins detected in samples depleted with the ProteoSpin™ or Pierce™ column. Displayed are a) the median total number of distinct proteins detected after three runs in samples depleted with the ProteoSpin™ (in black) or Pierce™ column (in grey) and b) the median number of proteins that were detected in all runs of samples depleted with the ProteoSpin™ (in black) or Pierce™ column (in grey). The minimum and maximum number of identifications are indicated by the error bars.

3.1.2 Retrieval Reliability of Abundance Variability

As biomarkers should allow for a stable detection, protein abundances should be detected with a variability below 20%. Data obtained from samples depleted with the ProteoSpin™ column showed to

be more consistent than data obtained from samples depleted with the Pierce™ column at all levels. As illustrated in Table 2 and Figure 8, the number of proteins for which the inter-prep variability was below 20% was higher in samples depleted with the ProteoSpin™ column compared to those depleted with the Pierce™ column. Similar to the total number of identifications, using the MaxLFQ algorithm decreased the number of proteins for which the inter-prep variability was below 20% in samples depleted with the ProteoSpin™ column. However, the opposite was true in samples depleted with the Pierce™ column. Statistics could not be performed to determine the significance of the differences in the number of proteins for which the inter-prep variability was below 20% because they were assessed using only one aliquot.

The number of proteins for which the inter-run variability was below 20% did not differ significantly between samples depleted with the ProteoSpin™ or Pierce™ column (p = 0,1875; Table 2). The use of the MaxLFQ algorithm increased this number in all samples (Table 2; Figure 9). This increase was borderline significant for samples depleted with the Pierce™ column (p = 0,0625), but not for samples depleted with the ProteoSpin™ column (p = 0,1250). When data was analysed using the MaxLFQ algorithm, the number of proteins for which the inter-run variability was below 20% did not differ between samples depleted with the ProteoSpin™ or Pierce™ column (p >0,9999; Table 2).

Another measure reflecting detection stability is the range of the inter-run variability per protein across different samples. The median range of the inter-run variability was higher in samples depleted with the ProteoSpin™ column (65,00% [21,30% – 124,80%]; Figure 10) compared to those depleted with the Pierce™ column (41,75% [17,14% – 114,50%]; Figure 10). This difference attenuated when the MaxLFQ algorithm was used (Figure 10). Moreover, the MaxLFQ algorithm decreased the median range of the inter-run variability across different samples for samples depleted with the ProteoSpin™ column (11,15% [0,69% – 49,22%]; Figure 10) and for samples depleted with the Pierce™ column (8,50% [1,77% – 79,34%]; Figure 10).

Table 2 Number of proteins detected with a CV below 20% at different levels.

	ProteoSpin™ column		Pierce™ column	
Variability	**MaxQuant**	**MaxLFQ**	**MaxQuant**	**MaxLFQ**
Inter-prep	124	116	66	98
Inter-run	95 [6 – 129]	119 [107 – 120]	23 [5 – 72]	117 [109 – 123]

Abbreviations: CV, coefficient of variation; prep, preparation

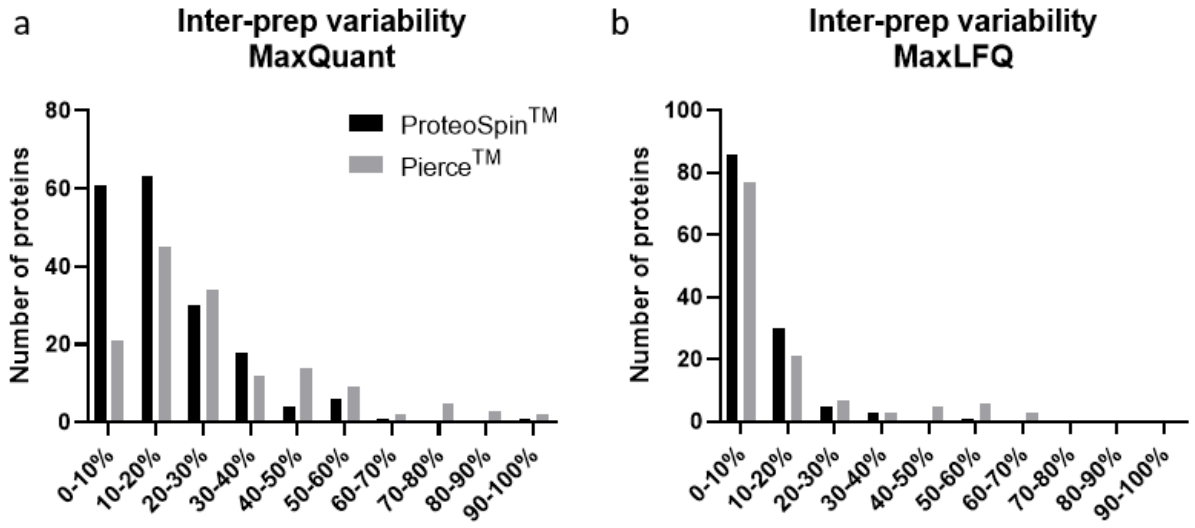

Figure 8 Distribution of the variation on the intensity of proteins detected in samples depleted with the ProteoSpin™ or Pierce™ column. The distribution of the inter-prep variability on the intensity of proteins detected in samples depleted with the ProteoSpin™ (in black) or Pierce™ column (in grey) is illustrated a) when data was analysed with the standard MaxQuant algorithm and b) the MaxLFQ algorithm.
Abbreviations: prep, preparation

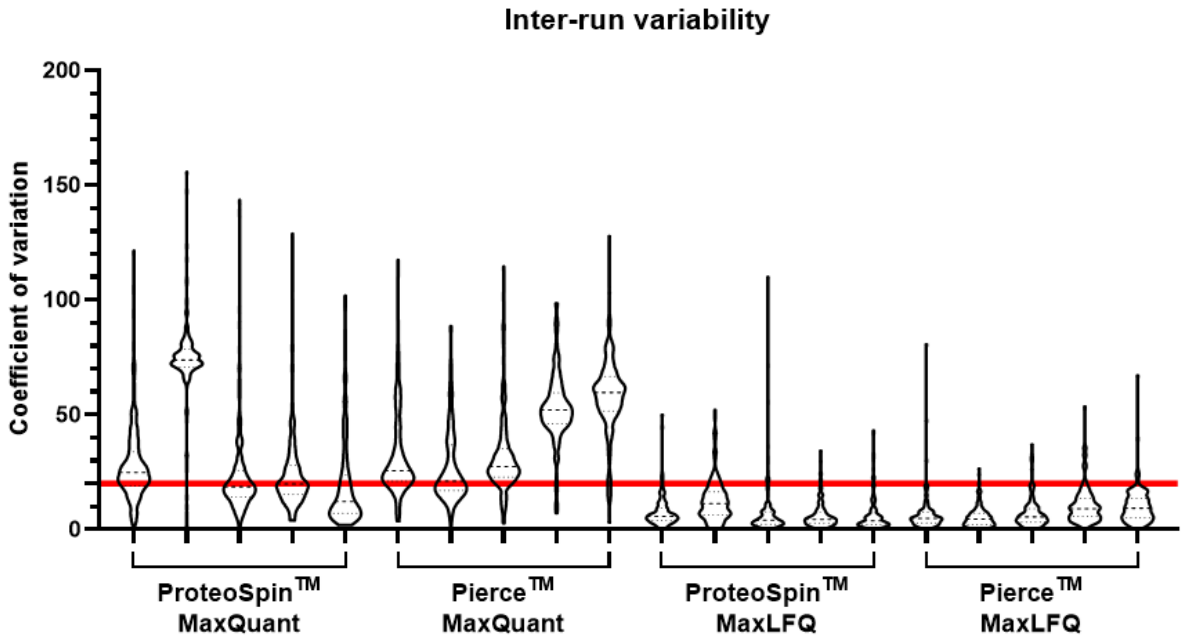

Figure 9 Distribution of the inter-run variability on the intensity of proteins detected in samples depleted with the ProteoSpin™ or Pierce™ column. The distribution of the inter-run variability on the intensity of detected proteins in samples depleted with the ProteoSpin™ or Pierce™ column is illustrated for each sample separately when data was analysed using the standard MaxQuant algorithm and the MaxLFQ algorithm. The red line indicates the 20% variation cut-off that was used in this thesis.
Legend: Dashed line, median; dotted lines, quartiles

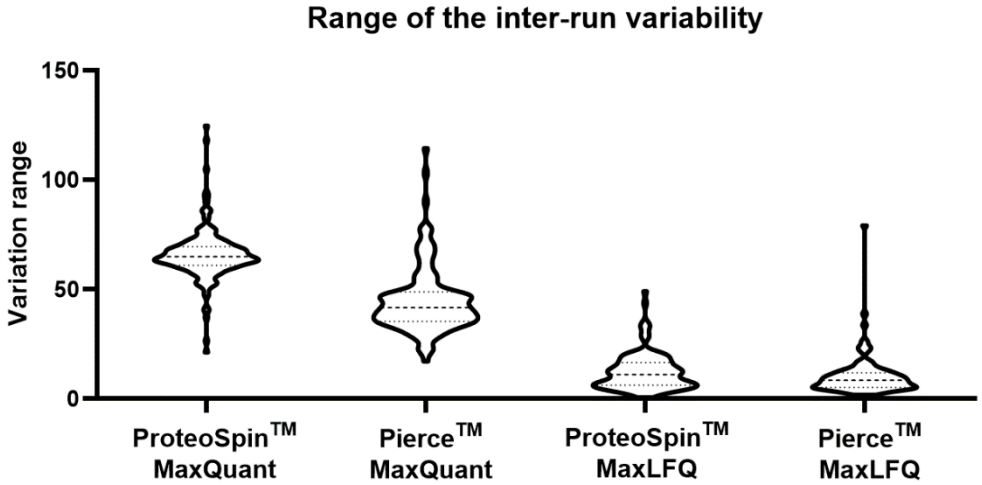

Figure 10 Distribution of the range of the inter-run variability on the intensity of proteins detected in samples depleted with the ProteoSpin™ or Pierce™ column. The distribution of the range of the inter-run variability on the intensity (the maximum CV subtracted with the minimum CV for each protein) of detected proteins in samples depleted with the ProteoSpin™ or Pierce™ column is illustrated when data was analysed using the standard MaxQuant algorithm and the MaxLFQ algorithm.
Legend: Dashed line, median; dotted lines, quartiles
Abbreviations: CV, coefficient of variation

3.2 Choice of Sample Source and Analytical Algorithm

3.2.1 Protein Yield

A first important measure to consider when comparing sample sources for biomarker discovery is the number of identifications. The median total number of distinct proteins retrieved after three runs was

significantly higher in PBMCs compared to plasma (950 [932 – 1029] vs 196 [191 – 198]; p = 0,0179; Figure 11a). This significant difference remained stable when the samples were analysed using the MaxLFQ algorithm (628 [579 – 974] vs 139 [133 – 168]; p = 0,0357; Figure 11a). Moreover, analysis with the MaxLFQ algorithm resulted in a significant decrease in the total number of identifications in plasma (p = 0,0079), but not in PBMCs (p = 0,400).

Similar to the optimisation phase for sample preparation of plasma, detection stability is pivotal when comparing different sample sources. The number of proteins identified in all runs of a sample was significantly higher in PBMCs compared to plasma (743 [705 – 903] vs 171 [165 – 174]; p = 0,0357; Figure 11b). This significant difference remained stable when data was analysed with the MaxLFQ algorithm (p = 0,0357, Figure 11b). Moreover, the MaxLFQ algorithm significantly decreased the number of proteins identified in all runs of a sample in plasma (116 [109 – 118]; p < 0,0079). This decrease was not significant in PBMCs (437 [404 – 553]; p = 0,1000).

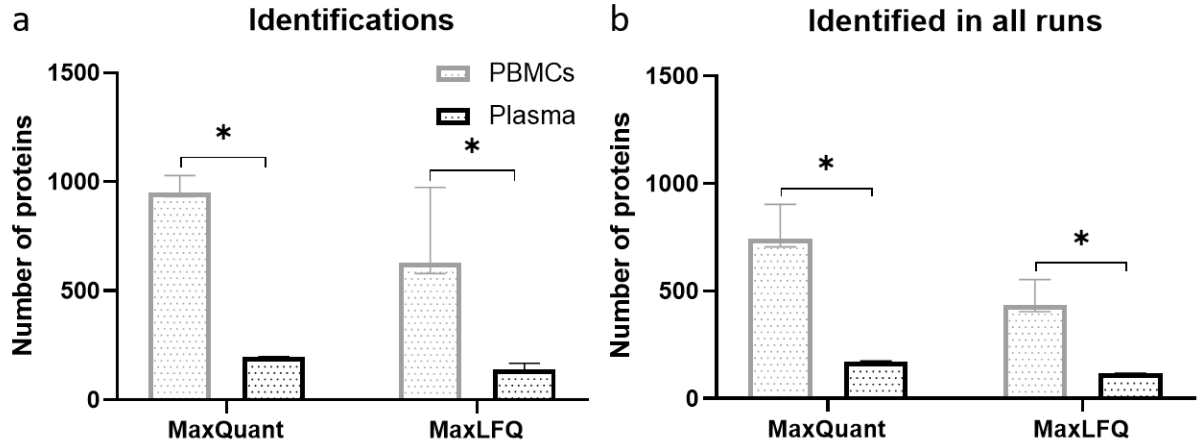

Figure 11 Number of proteins detected in PBMCs using the standard protocol and plasma using the optimised protocol. Displayed are a) the median total number of distinct proteins detected after three runs in plasma samples prepped with the optimised protocol (in black) and PBMC samples prepped with the standard protocol (in grey) when samples were analysed with the standard MaxQuant algorithm or the MaxLFQ algorithm and b) the median number of proteins that were detected in all runs of plasma samples prepped with the optimised protocol (in black) and PBMC samples prepped with the standard protocol (in grey). The minimum and maximum number of identifications are indicated by the error bars.
Legend: *: p-value ≤ 0,05

3.2.2 Retrieval Reliability of Abundance Variability

The variability on detected protein abundances was calculated to assess which sample source would yield the most reliable results. As can be concluded from Figure 12, the overall variation is larger in PBMCs compared to plasma (depleted with the ProteoSpin™). However, PBMCs generated a higher number of proteins with an inter-run variability below 20% than plasma (p = 0,2500; Table 3). Moreover, MaxLFQ positively influenced the number of proteins identified in PBMCs for which the inter-run variability was below 20% (p = 0,4000; Table 3; Figure 12), similar to plasma samples (p = 0,1349; Table 3; Figures 9 and 12). This number remained larger in PBMCs compared to plasma (p = 0,0179; Table 3). Also the number of proteins identified in PBMCs for which the inter-individual variability was below 20% increased when the MaxLFQ algorithm was used (Table 3; Figure 13). Unfortunately, the preliminary data was not sufficient to assess all types of variability. Therefore, the inter-prep and inter-individual variability could not be determined for PBMCs and plasma, respectively.

The median range of the inter-run variability was higher in plasma samples (65,04% [21,30% – 124,80%]) compared to PBMCs (30,41% [0,15% – 146,30%]; Figure 14). This difference remained stable

when data was analysed with MaxLFQ (Figure 14). Moreover, MaxLFQ decreased the median range of the inter-run variability across different samples in PBMCs (12,61% [0,43% – 96,84%]; Figure 14) and plasma (11,15% [0,69% – 49,22%]; Figure 14).

Table 3 Number of proteins detected with a CV below 20% at different levels.

	PBMCs		Plasma	
Variability	**MaxQuant**	**MaxLFQ**	**MaxQuant**	**MaxLFQ**
Inter-prep	NA	NA	124	116
Inter-run	277 [75 – 439]	433 [323 – 561]*	95 [6 – 129]	119 [107 – 120]*
Inter-individual	153	248	NA	NA

Legend: NA, not applicable; *, p-value < 0,05
Abbreviations: CV, coefficient of variation; PBMCs, peripheral blood mononuclear cells; prep, preparation

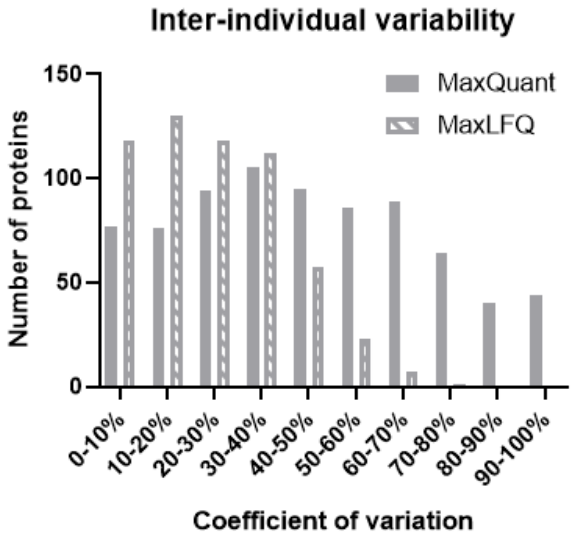

Figure 13 Distribution of the inter-individual variability on the intensity of proteins detected in PBMCs when analysing data with MaxQuant and MaxLFQ. The distribution of the inter-individual variability on the intensity of proteins detected in PBMCs is illustrated when data is analysed using the standard MaxQuant algorithm (in grey) and the MaxLFQ algorithm (striped pattern).
Abbreviations: PBMCs, peripheral blood mononuclear cells

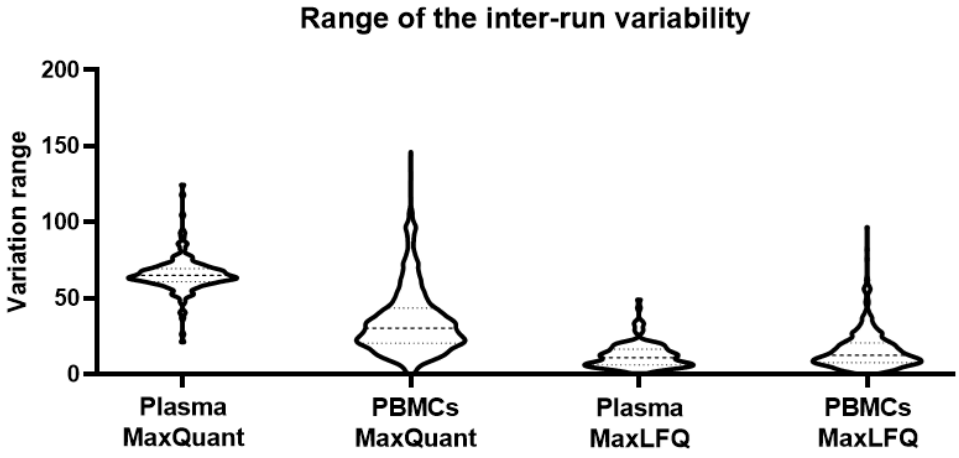

Figure 14 Distribution of the range of the inter-run variability on the intensity of proteins detected in PBMCs and plasma. The distribution of the range of the inter-run variability on the intensity (the maximum CV subtracted with the minimum CV for each protein) of detected proteins in PBMCs and plasma is illustrated when data was analysed using the standard MaxQuant algorithm and the MaxLFQ algorithm.
Legend: Dashed line, median; dotted lines, quartiles
Abbreviations: CV, coefficient of variation

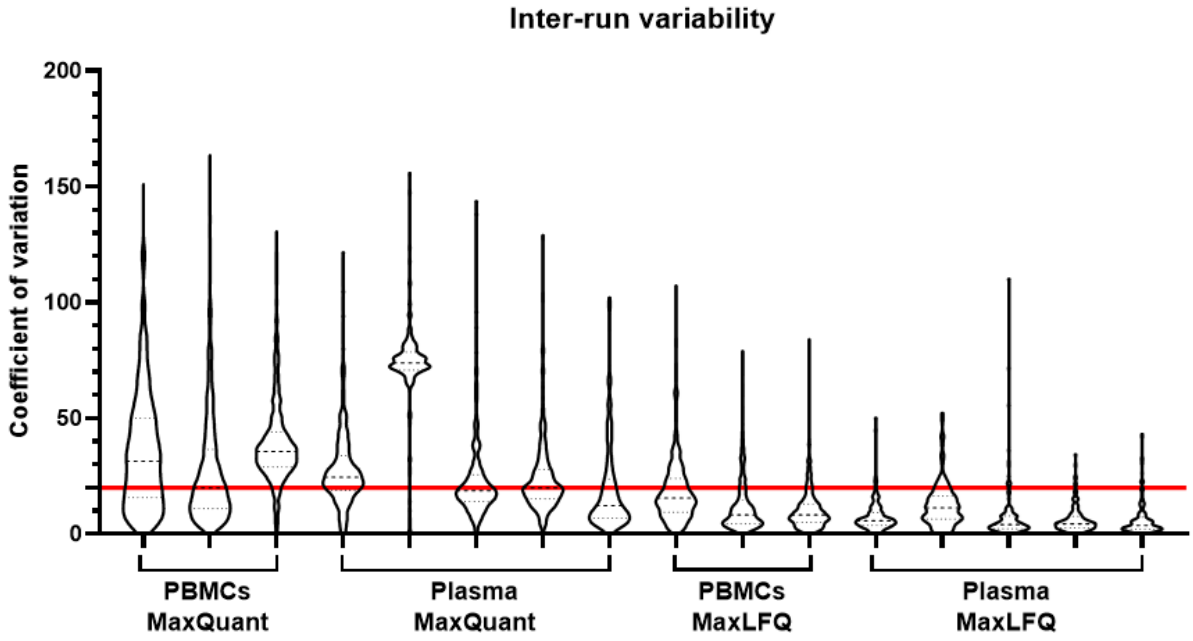

Figure 12 Distribution of the inter-run variability on the intensity of proteins detected in plasma and PBMCs. The distribution of the inter-run variability on the intensity of proteins detected in plasma and plasma is illustrated for each sample separately when data was analysed using the standard MaxQuant algorithm and the MaxLFQ algorithm. The red line indicates the 20% variation cut-off that was used in this thesis.
Legend: Dashed line, median; dotted lines, quartiles
Abbreviations: PBMCs, peripheral blood mononuclear cells

3.3 Effect of the Length of the LC Gradient on the Number of Identifications in Plasma

Plasma samples that were depleted with the ProteoSpin™ column were used to test variable LC gradient lengths and data was analysed using the standard MaxQuant algorithm. Although a small decrease in the number of identifications was observed when the gradient was increased from 155 to 180 minutes (186 vs 160), this number increased again when the gradient was lengthened to 240, 300, 360 and 440 minutes (233, 219, 236, 235, respectively; Figure 15). Therefore, there seems to be a positive effect of the length of the LC gradient on the number of identifications, reaching a maximum at 240 minutes. As each data point of this experiment was obtained from only one replicate of one sample, significant differences could not be determined. Moreover, the effect of the length of the LC gradient on the number of identifications in PBMCs could not be assessed due to the COVID-19 lockdown.

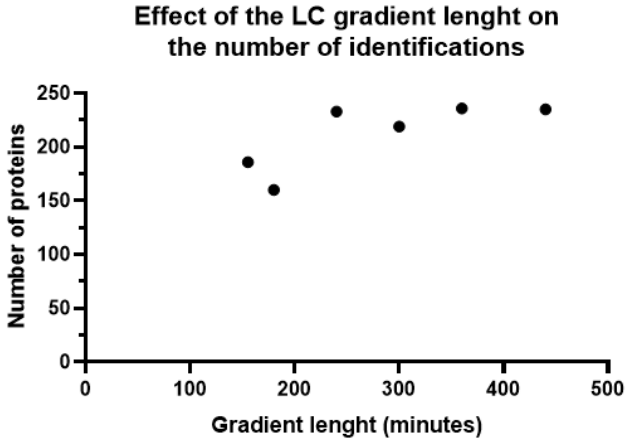

Figure 15 Effect of the gradient length on the number of identifications in plasma. Displayed are the number of proteins that are identified in plasma samples dependent on the length of the LC gradient.
Abbreviations: LC, liquid chromatography

3.4 Demographics

Group-mean differences were not significant for age, gender, BMI and smoking. Moreover, HDRS scores indicate that all patients showed more depressive symptoms than HCs. More specifically, MDD and BD-D patients had the highest HDRS scores. The positive scale of the PANSS indicated that psychotic symptoms were only present in SZ patients. The same patient group scored highest on the YMRS for manic symptoms. The YMRS score of SZ and BD-M patients differed significantly from those of MDD patients, BD-D patients and healthy controls. Demographic variables and symptom scores are presented in Table 4.

Table 4 Demographics of MDD, BD-D, BD-M and SZ patients and healthy controls.

	HC	MDD	BD-D	BD-M	SZ
N	6	5	3	4	4
Age	32 ± 8,56	28,6 ± 10,95	24,67 ± 3,06	36,5 ± 11,5	34,5 ± 11,5
Gender (M/F)	2/4	1/4	1/2	1/3	1/3
BMI	25,9 ± 3,01	24,74 ± 4,4	24,67 ± 2,84	31,42 ± 2,8	25,68 ± 6,14
Smoking	3	1	2	2	3
HDRS	1 ± 1,55	20,2 ± 1,79*	21 ± 3,46*	10 ± 6,68	14 ± 4,32*
PANSS-P	7 ± 0	7 ± 0	7 ± 0	7 ± 0	20,25 ± 4,03*
YMRS	0,17 ± 0,41	1,6 ± 1,52	2,67 ± 2,31	13,5 ± 0,58*	18,25 ± 1,5*

Data are presented as mean ± SD; * $p < 0.05$; significantly different from healthy controls.
Abbreviations: BD-D, bipolar patients in a depressive state; BD-M, bipolar patients in a manic state; BMI, body mass index; F, female; HC, healthy controls; HDRS, Hamilton Depression Rating Scale; M, male; MDD, major depressive disorder patients; n, sample size; PANSS-P, positive scale of the Positive And Negative Syndrome Scale; SZ, schizophrenia patients, YMRS, Young Mania Rating Scale

3.5 Identification of Biomarker Candidates for Differential Diagnosis

In total, 4271 proteins were identified, of which 2651 were detected in each sample and thus selected for further analysis. Log_2 fold change expression ratios compared to HCs were calculated for each experimental group. The highest number of differentially expressed proteins (DEPs) were identified in BD-D patients (141 DEPs) followed by BD-M patients (138 DEPs), MDD patients (125 DEPs) and SZ patients (92 DEPs). When comparing patient groups with each other, 133 proteins were differentially expressed between MDD and BD-D patients while 115 proteins were differentially expressed between BD-M and SZ patients.

DEPs were only included in the lists of biomarker candidates when its log_2 fold change expression ratio was larger than two SD to i) the patient/patient mean log_2 fold change expression ratio and ii) the patient/HC mean log_2 fold change expression ratio in at least one of the two patient groups. This *modus operandi* resulted in the identification of 67 biomarker candidates for differential diagnosis of MDD and BD-D patients (Table 5) and 78 biomarker candidates for differential diagnosis of BD-M and SZ patients (Table 6), as illustrated by the red outlines in Figure 16. Biomarkers candidates that are differentially expressed in all three comparisons are indicated by the black outlines in Figure 16. The biomarker candidates lists have 24 biomarker candidates in common.

Interestingly, two of the biomarker candidates for differential diagnosis of BD-M and SZ patients (Apolipoprotein C-III (APOC3) and HLA class I histocompatibility antigen, B-7 alpha chain (HLA-B7)) were found to be contra-regulated in these pathologies. APOC3 showed to be down-regulated in BD-M patients (-0.24 fold change in expression from HCs) and up-regulated in SZ patients (0.23 fold change

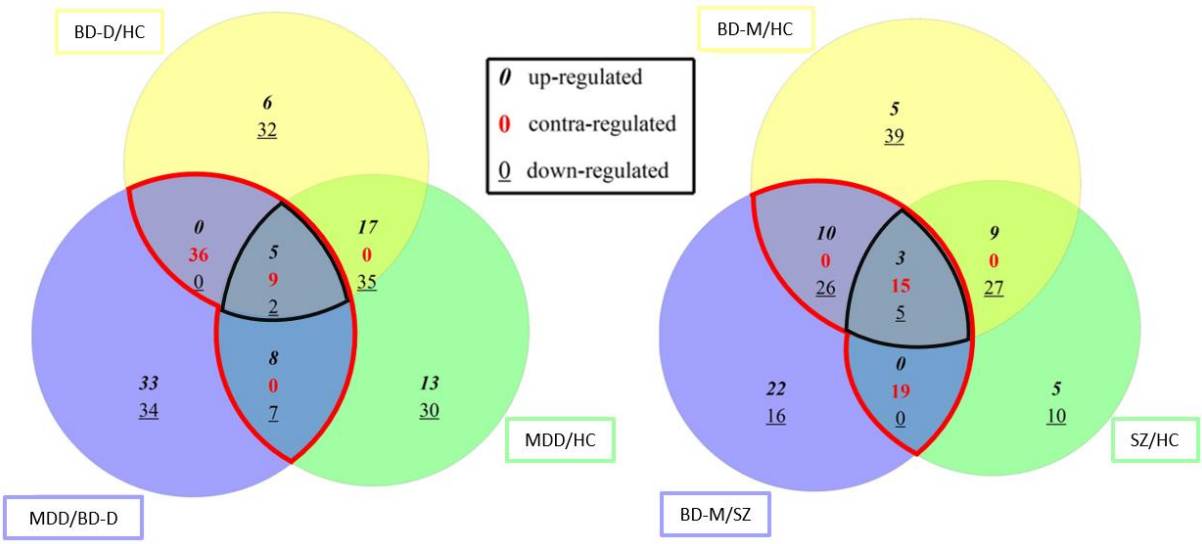

Figure 16 Venn diagrams representing the number DEPs identified in PBMCs of MDD, BD-D, BD-M and SZ patients and their overlap. Each circle contains the number of proteins that were differentially expressed between the indicated experimental groups. Moreover, the diameter of each circle is correlated with this number (diameter increases when the number of DEPs in that circle is larger). The numbers in overlapping areas of the circles represent the numbers of proteins that were differentially expressed in those comparisons. DEPs that were included in the lists of biomarker candidates are indicated by the red outlines. The black outlines mark proteins that are differentially expressed in all comparisons.
Abbreviations: BD-D, bipolar patients in a depressive state; BD-M, bipolar patients in a manic state; DEPs, differentially expressed proteins; HC, healthy controls; MDD, major depressive disorder; PBMCs, peripheral blood mononuclear cells; SZ, schizophrenia

in expression from HCs) while HLA-B was down-regulated in SZ patients (-0.32 fold change in expression from HCs) and up-regulated in BD-M patients (0.35 fold change in expression from HCs). Although not one biomarker candidate for differential diagnosis of MDD and BD-D patients was contra-regulated in these experimental groups, 16 biomarkers candidates showed a significant $\log_2$ fold change in expression for both experimental groups in the same direction (i.e. either up- or down-regulated). This was also the case for 21 biomarker candidates for differential diagnosis of BD-M and SZ patients. Even more interestingly, 7 of the 24 overlapping biomarker candidates (Cathelicidin antimicrobial peptide; Granzyme H; Granulysin (Fragment); Lactotransferrin; Bactericidal permeability-increasing protein; Neutrophil gelatinase-associated lipocalin; Keratin, type I cytoskeletal 9) were differentially expressed in all patient groups compared to controls. Four of these proteins were up-regulated and three were down-regulated in all patient groups.

Table 5 List of biomarker candidates ranked according to the largest discriminatory potential between MDD and BD-D patients.

Protein	MDD/BD-D 2SD = 0,1647	MDD/HC 2SD = 0,1797	BD-D/HC 2SD = 0,2173
HLA class I histocompatibility antigen, A-24 alpha chain	0,6365	0,0769	-0,5595
Keratin, type II cytoskeletal 1	0,6253	-0,1888	-0,8141
HLA class I histocompatibility antigen, B-18 alpha chain	0,6217	-0,0052	-0,6269
Keratin, type I cytoskeletal 9	0,5759	-0,2284	-0,8043
HLA class I histocompatibility antigen, A-2 alpha chain	0,4704	0,0306	-0,4398
Histone H2A type 1	0,4620	-0,4962	-0,9582
Keratin, type I cytoskeletal 10	0,4321	-0,0968	-0,5288
High mobility group nucleosome-binding domain-containing protein 4	0,3720	0,0557	-0,3163
HD domain-containing protein 2	0,3667	-0,0306	-0,3973
Galectin-10	0,3286	0,3278	-0,0008

Beta-hexosaminidase	0,3241	-0,2927	-0,6168
Hemoglobin subunit gamma-1	0,3038	0,5891	0,2853
HLA class I histocompatibility antigen, Cw-1 alpha chain	0,2943	0,3098	0,0155
Keratin, type II cytoskeletal 2 epidermal	0,2553	-0,1003	-0,3556
40S ribosomal protein S28	0,2534	0,2547	0,0013
Histone H2A type 2-C	0,2533	0,0337	-0,2196
Neutrophil gelatinase-associated lipocalin	0,2530	0,5785	0,3255
OCIA domain-containing protein 2	0,2491	-0,3010	-0,5501
DNA dC->dU-editing enzyme APOBEC-3C	0,2472	0,1871	-0,0601
Mitogen-activated protein kinase 13	0,2445	-0,0001	-0,2446
Histone H1.4	0,2417	0,2590	0,0173
Parathymosin	0,2413	0,0103	-0,2309
Hemoglobin subunit gamma-2	0,2333	0,3311	0,0977
PHD finger protein 6	0,2303	-0,1533	-0,3836
Protein IWS1 homolog	0,2299	-0,1645	-0,3943
Peptidyl-prolyl cis-trans isomerase G	0,2244	-0,0559	-0,2804
39S ribosomal protein L28, mitochondrial	0,2243	0,0025	-0,2218
Carcinoembryonic antigen-related cell adhesion molecule 8	0,2204	0,2657	0,0453
Alpha-ketoglutarate-dependent dioxygenase FTO	0,2180	-0,0655	-0,2835
Bactericidal permeability-increasing protein	0,2172	0,5281	0,3109
Lactotransferrin	0,2093	0,5972	0,3878
RNA-binding protein 42	0,2093	-0,0229	-0,2322
N-sulphoglucosamine sulphohydrolase	0,2063	-0,0455	-0,2519
Cytochrome b-c1 complex subunit 9	0,2038	-0,2341	-0,4379
ADP-ribosylation factor-like protein 8B	0,2020	0,3181	0,1160
Granzyme K	0,1999	-0,0261	-0,2260
Cathelicidin antimicrobial peptide	0,1993	0,7006	0,5013
Macrophage migration inhibitory factor	0,1974	-0,1647	-0,3621
Protein LSM14 homolog A	0,1955	-0,0557	-0,2512
Histone H2B type 3-B	0,1919	-0,1083	-0,3002
Liver carboxylesterase 1	0,1916	-0,1151	-0,3067
E3 ubiquitin-protein ligase RNF123	0,1888	-0,0515	-0,2403
ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 2	0,1885	-0,0893	-0,2779
WD repeat-containing protein 43	0,1755	-0,2546	-0,4301
Solute carrier family 35 member F6	0,1752	-0,0732	-0,2484
Mitochondrial-processing peptidase subunit alpha	0,1734	-0,1452	-0,3187
Granzyme H	0,1708	-0,4005	-0,5713
Transcription factor BTF3 homolog 4	0,1707	-0,0557	-0,2264
Tyrosine-protein phosphatase non-receptor type 7 (Fragment)	0,1704	-0,1009	-0,2712
Non-histone chromosomal protein HMG-17	0,1685	-0,1175	-0,2860
Granulysin (Fragment)	0,1653	-0,3879	-0,5531
Sulfotransferase 1A1	-0,1747	-0,2174	-0,0427
Prenylcysteine oxidase-like	-0,1783	0,0939	0,2721
GTPase IMAP family member 5	-0,1788	-0,2275	-0,0487
Platelet glycoprotein VI	-0,2016	-0,2013	0,0003
Endoplasmic reticulum aminopeptidase 2	-0,2115	-0,4551	-0,2437
Tropomyosin beta chain	-0,2296	0,0154	0,2450

Trem-like transcript 1 protein	-0,2369	-0,6264	-0,3895
Tubulin beta-3 chain	-0,2431	-0,3591	-0,1160
HLA class I histocompatibility antigen, B-41 alpha chain	-0,2634	0,0285	0,2919
Bardet-Biedl syndrome 12 protein	-0,2769	-0,2607	0,0163
HLA class I histocompatibility antigen, A-68 alpha chain	-0,3004	-0,0024	0,2980
cGMP-inhibited 3',5'-cyclic phosphodiesterase B	-0,3542	-0,3100	0,0442
Retinoblastoma-like protein 1	-0,3908	0,1767	0,5675
Plexin-A4	-0,3922	-0,2250	0,1671
Interferon-induced GTP-binding protein Mx1	-0,4527	-0,1016	0,3511
HLA class II histocompatibility antigen, DRB1-16 beta chain	-0,5935	-0,0193	0,5742

Legend: red, increased expression in pathology compared to HCs; green, decreased expression in pathology compared to HCs; Abbreviations: BD-D, bipolar patients in a depressive state; HC, healthy controls; MDD, patients with major depressive disorder

Table 6 List of biomarker candidates ranked according to the largest discriminatory potential between BD-M and SZ patients.

Protein	**BD-M/SZ** 2SD = 0,1585	**BD-M/HC** 2SD = 0,1925	**SZ/HC** 2SD = 0,1922
HLA class I histocompatibility antigen, B-7 alpha chain	0,6693	0,3480	-0,3213
HLA class I histocompatibility antigen, A-34 alpha chain	0,5097	0,4055	-0,1042
Alanine--tRNA ligase, mitochondrial	0,4497	0,4428	-0,0069
40S ribosomal protein S21	0,2967	0,2236	-0,0731
Neutrophil defensin 1	0,2924	0,9423	0,6499
C-C motif chemokine 5	0,2824	0,3363	0,0539
Calumenin	0,2757	0,2655	-0,0102
40S ribosomal protein S28	0,2623	0,2177	-0,0446
Myristoylated alanine-rich C-kinase substrate	0,2313	0,2657	0,0344
Pre-mRNA-splicing factor SYF1	0,2067	0,2137	0,0070
Aldo-keto reductase family 1 member C3	0,2003	-0,0080	-0,2083
Granzyme H	0,2001	-0,2364	-0,4365
Platelet basic protein	0,1960	0,2282	0,0323
Ubiquitin-conjugating enzyme E2 D1	0,1940	0,2290	0,0350
Macrophage migration inhibitory factor	0,1689	-0,1112	-0,2802
H(+)/Cl(-) exchange transporter 3	0,1669	0,4618	0,2949
Cathepsin G	0,1620	0,5093	0,3473
Solute carrier family 2, facilitated glucose transporter member 1	-0,1615	0,0950	0,2565
Kynurenine--oxoglutarate transaminase 3	-0,1630	-0,2009	-0,0379
E3 ubiquitin-protein ligase TRIM22	-0,1726	-0,2934	-0,1209
Peroxiredoxin-2	-0,1727	0,0229	0,1956
Protein S100-A8	-0,1733	0,4675	0,6407
Mitochondrial carrier homolog 1 (Fragment)	-0,1743	-0,3177	-0,1434
Eukaryotic elongation factor 2 kinase	-0,1743	-0,6258	-0,4515
COMM domain-containing protein 4	-0,1792	-0,3078	-0,1286
Haptoglobin	-0,1798	0,0273	0,2071
Lysosomal alpha-glucosidase	-0,1813	-0,3132	-0,1319
Carbonic anhydrase 1	-0,1818	0,0233	0,2051
Apolipoprotein A-I	-0,1846	-0,2726	-0,0880
Ubiquinone biosynthesis monooxygenase COQ6, mitochondrial	-0,1911	-0,2822	-0,0911
Small nuclear ribonucleoprotein E	-0,1939	-0,2548	-0,0610

Keratin, type I cytoskeletal 9	-0,1973	-0,4589	-0,2616
Cytochrome b-c1 complex subunit 9	-0,1973	-0,3178	-0,1205
Condensin complex subunit 1	-0,1991	-0,3029	-0,1038
HLA class I histocompatibility antigen, alpha chain E	-0,2041	-0,3558	-0,1518
Protein S100-P	-0,2061	0,3571	0,5631
HLA class I histocompatibility antigen, A-24 alpha chain	-0,2084	-0,3915	-0,1831
HLA class I histocompatibility antigen, A-2 alpha chain	-0,2151	-0,2707	-0,0556
Putative beta-actin-like protein 3	-0,2194	-1,0643	-0,8449
Carcinoembryonic antigen-related cell adhesion molecule 8	-0,2202	0,3197	0,5399
CAP-Gly domain-containing linker protein 2	-0,2228	-0,1947	0,0281
Band 3 anion transport protein	-0,2231	-0,0006	0,2225
Nicotinamide phosphoribosyltransferase	-0,2270	-0,0183	0,2086
Arylsulfatase B	-0,2294	-0,2125	0,0169
Phospholipase D3	-0,2538	-0,2402	0,0136
Bactericidal permeability-increasing protein	-0,2559	0,4348	0,6907
Zinc finger protein 648	-0,2593	0,2296	0,4889
Phosphomevalonate kinase	-0,2640	-0,0555	0,2085
Rap guanine nucleotide exchange factor 3	-0,2687	-0,2445	0,0243
cGMP-inhibited 3',5'-cyclic phosphodiesterase B	-0,2699	-0,2781	-0,0081
AP2-associated protein kinase 1	-0,2741	-0,2967	-0,0226
Histone H2A type 2-C	-0,2773	-0,2280	0,0493
Lactotransferrin	-0,2825	0,5031	0,7856
Gamma-tubulin complex component 2	-0,2828	-0,0728	0,2100
Protein Mpv17	-0,2880	-0,5827	-0,2947
Apolipoprotein C-I	-0,2884	-0,3912	-0,1028
HLA class I histocompatibility antigen, B-18 alpha chain	-0,2895	-0,4689	-0,1794
Hemoglobin subunit beta	-0,2896	0,2293	0,5189
Hemoglobin subunit alpha	-0,3023	0,1906	0,4928
Tubulin beta-3 chain	-0,3033	-0,3280	-0,0247
AP-3 complex subunit sigma-1	-0,3067	0,0331	0,3398
Non-histone chromosomal protein HMG-17	-0,3244	-0,1968	0,1276
Catechol O-methyltransferase	-0,3259	-0,0974	0,2285
Zinc finger ZZ-type and EF-hand domain-containing protein 1	-0,3411	-0,0943	0,2468
Granulysin (Fragment)	-0,3436	-0,5838	-0,2403
Sulfotransferase 1A1	-0,3454	-0,2381	0,1073
Selenium-binding protein 1	-0,3492	-0,0419	0,3073
Bardet-Biedl syndrome 12 protein	-0,3760	-0,1611	0,2149
Cathelicidin antimicrobial peptide	-0,4017	0,5550	0,9567
Protein S100-A12	-0,4063	0,2271	0,6333
Matrix metalloproteinase-9	-0,4075	0,3411	0,7487
Retinoblastoma-like protein 1	-0,4267	0,3676	0,7944
Hemoglobin subunit gamma-2	-0,4323	0,0840	0,5162
Neutrophil gelatinase-associated lipocalin	-0,4469	0,4932	0,9401
Apolipoprotein C-III	-0,4737	-0,2457	0,2281
HLA class II histocompatibility antigen, DRB1-11 beta chain	-0,5360	-0,5064	0,0296
Galectin-10	-0,6977	-0,1375	0,5602
Hemoglobin subunit gamma-1	-0,7167	0,1511	0,8678

Legend: red, increased expression in pathology compared to HCs; green, decreased expression in pathology compared to HCs;
Abbreviations: BD-M, bipolar patients in a manic state; HC, healthy controls; SZ, schizophrenia patients

3.6 Pathway Analysis

Pathway analysis was performed to identify biological functions that are differentially affected between the pathologies of interest. A gene set was considered to be significantly enriched when the q-value was below 0,05. 143 and 307 significantly enriched gene sets were identified in MDD or BDD patients compared to HCs, respectively (data not shown). When comparing expression values of MDD and BD-D patients with each other, 499 gene set were significantly enriched (data not shown). 139 of those were also enriched when expression values of either MDD or BD-D patients were compared to HCs (Table S2 in Supplementary materials). Interactions between these pathways are visualised in Figure 17. 39 gene sets did not show any interactions with other significantly enriched gene sets (i.e. single nodes) and were therefore not included in Figure 17.

After correction for multiple testing, no gene sets were identified as being significantly enriched when expression values of BD-M and SZ patients were compared. In case BD-M or SZ patients were compared to HCs, 63 and 76 gene sets were enriched, respectively (data not shown).

3.7 Network Analysis

Protein-protein interactions between biomarker candidates for differential diagnosis of MDD and BD-D patients (Figure 18), on the one hand, and BD-M and SZ patients (Figure 19), on the other, were generated via the STRING database. Network characteristics are summarised in Table 7. Hubs and bottlenecks of both networks are displayed in Table 8. The protein-protein interaction network of biomarker candidates for differential diagnosis of MDD and BD-D patients consists of four hub-bottlenecks. This is the case for 10 biomarker candidates for differential diagnosis of BD-M and SZ patients, of which none overlapped with those for differential diagnosis of MDD and BD-D patients. Hub-bottleneck proteins are highlighted in blue-greenish in Table 8.

These networks were also subjected to cluster analysis, which resulted in the identification of seven clusters in both networks. Subsequently, functional enrichment analysis was performed on each of these clusters separately. The results of these analyses as well as the characteristics of the clusters are summarised in Table 9 for biomarker candidates for differential diagnosis of MDD and BD-D patients and Table 10 for biomarker candidates for differential diagnosis of BD-M and SZ patients.

Table 7 Characteristics of the protein-protein interaction network based on biomarker candidates for differential diagnosis.

	MDD/BD-D	BD-M/SZ
Number of nodes	58	70
Number of edges	35	118
Isolated components	23	20
Highest degree	4	14
Average number of neighbours	1,207	3,086
Cluster coefficient	0,230	0,297
Network density	0,021	0,045

Abbreviations: BD-D, bipolar patients in a depressive state; BD-M, bipolar patients in a manic state; MDD, major depressive disorder; SZ, schizophrenia

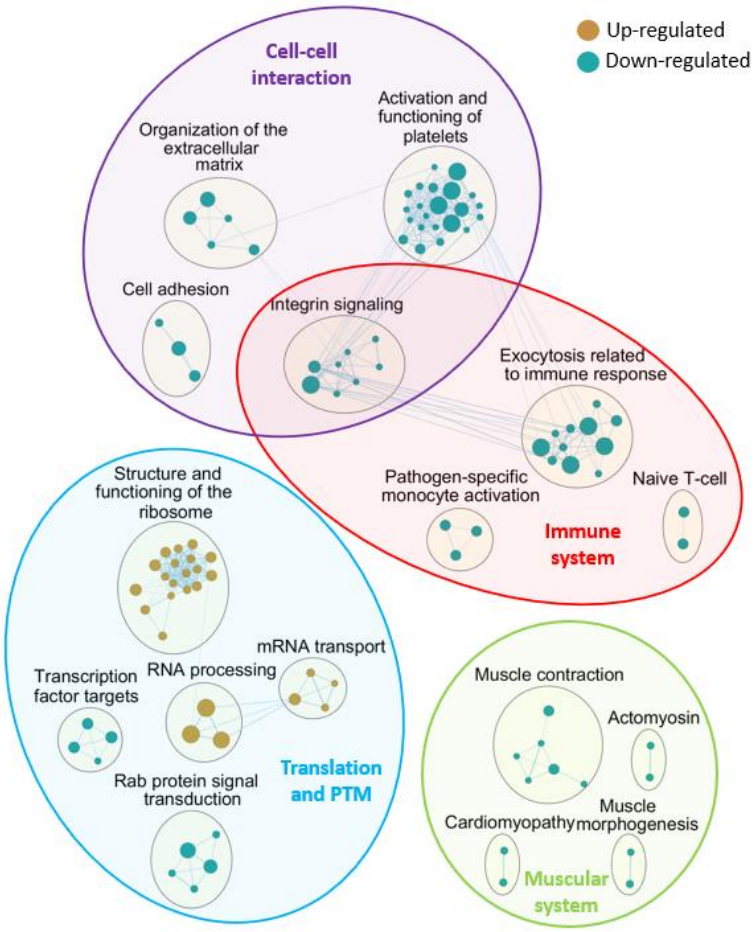

Figure 17 Interaction network of gene sets that are differentially enriched between MDD and BD-D patients. This network visualises the interactions (edges) between significantly enriched gene sets (nodes) when MDD and BD-D patients are compared. The size of each node is in function of the size of the gene set (larger diameter with increasing gene set size) and the node colour corresponds to the direction in which the gene sets are regulated. The width of edges corresponds to the number of overlapping genes within the gene sets (wider edge with increasing number of overlapping genes).
Abbreviations: BD-D, bipolar patients in a depressive state; MDD, major depressive disorder; PTM, post-translation modifications

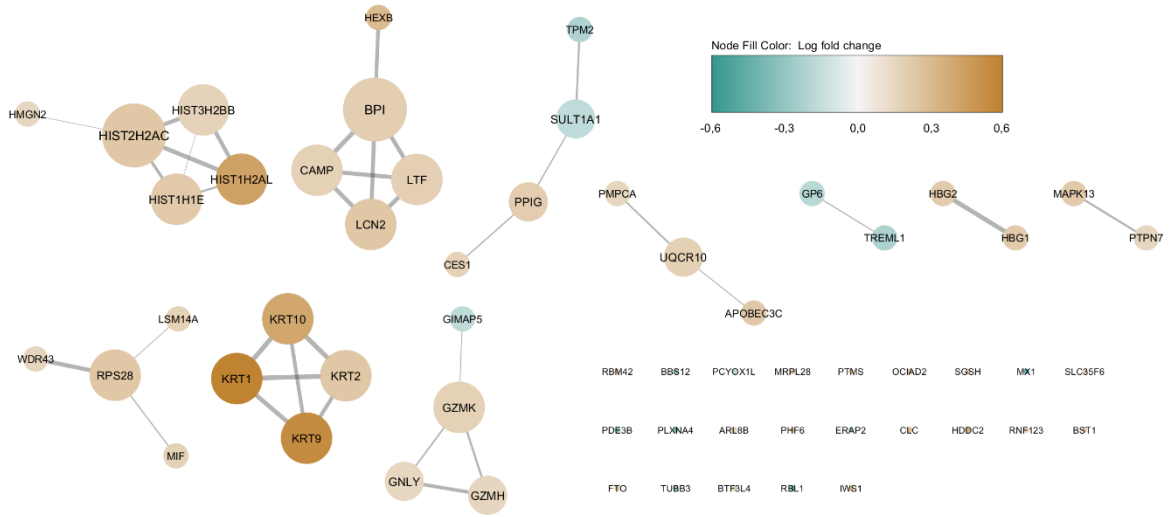

Figure 18 Protein-protein interaction network based on biomarker candidates for differential diagnosis of MDD and BD-D patients. This network visualises the interactions (edges) between biomarker candidates for differential diagnosis of MDD and BD-D patients (nodes). The name of each node corresponds to the gene from which the biomarker candidate is derived. Moreover, the size of each node is in function of its degree (larger diameter with increasing degree) and the node colour corresponds to the $\log_2$ fold change expression ratio of that protein when comparing MDD to BD-D patients. The width of edges corresponds to the STRING confidence score (wider edge with increasing confidence score).
Abbreviations: BD-D, bipolar patients in a depressive state; MDD, major depressive disorder

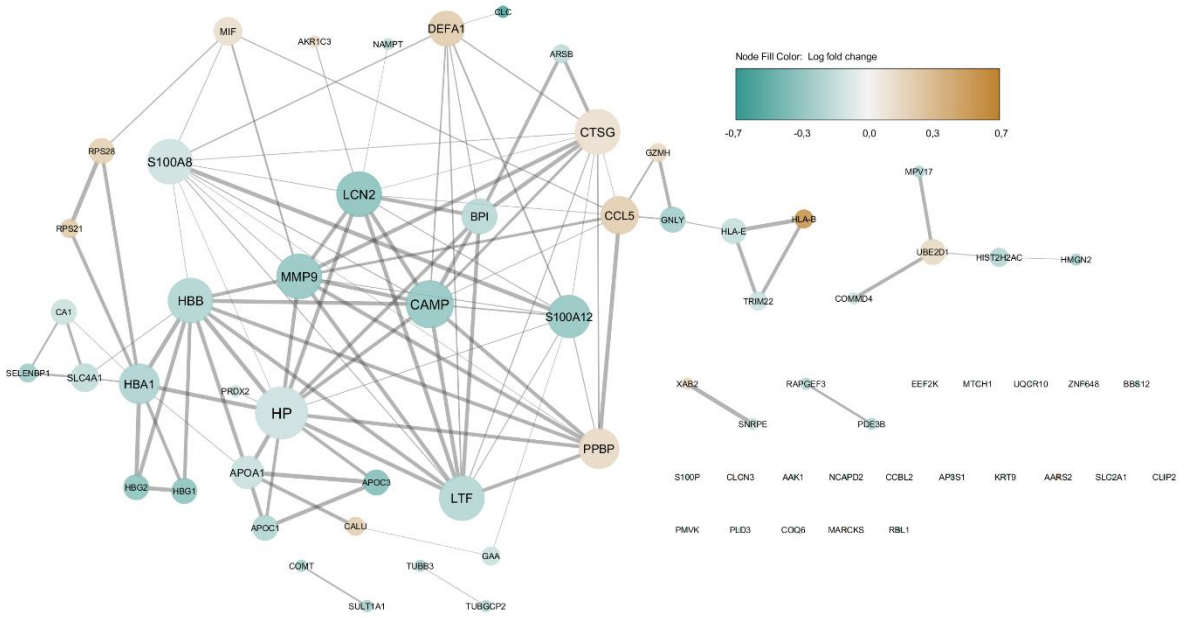

Figure 19 Protein-protein interaction network based on biomarker candidates for differential diagnosis of BD-M and SZ patients. This network visualises the interactions (edges) between biomarker candidates for differential diagnosis of BD-M and SZ patients (nodes). The name of each node corresponds to the gene from which the biomarker candidate is derived. Moreover, the size of each node is in function of its degree (larger diameter with increasing degree) and the node colour corresponds to the $\log_2$ fold change expression ratio of that protein when comparing BD-M to SZ patients. The width of edges corresponds to the STRING confidence score (wider edge with increasing confidence score).
Abbreviations: BD-M, bipolar patients in a manic state; SZ, schizophrenia

Table 8 Top 20% hubs and bottlenecks of both networks containing biomarker candidates for differential diagnosis.

	MDD/BD-D		BD-M/SZ	
Rank	**Hubs**	**Bottlenecks**	**Hubs**	**Bottlenecks**
1	Histone H2A type 2-C	Histone H2A type 2-C	Haptoglobin	C-C motif chemokine 5
2	Bactericidal permeability-increasing protein	Bactericidal permeability-increasing protein	Cathelicidin antimicrobial peptide	Haptoglobin
3	Histone H2B type 3-B	40S ribosomal protein S28	Neutrophil gelatinase-associated lipocalin	Hemoglobin subunit beta
4	Keratin, type II cytoskeletal 2 epidermal	Granzyme K	Protein S100-A8	Granulysin
5	Lactotransferrin	Sulfotransferase 1A1	Matrix metalloproteinase-9	Hemoglobin subunit alpha
6	Histone H2A type 1	Peptidyl-prolyl cis-trans isomerase G	Cathepsin G	Neutrophil gelatinase-associated lipocalin
7	Keratin, type I cytoskeletal 10	Cytochrome b-c1 complex subunit 9	Lactotransferrin	HLA class I histocompatibility antigen, alpha chain E
8	Keratin, type I cytoskeletal 9	Histone H2B type 3-B	Hemoglobin subunit beta	Cathelicidin antimicrobial peptide
9	Histone H1.4	Bardet-Biedl syndrome 12 protein	Protein S100-A12	Matrix metalloproteinase-9
10	40S ribosomal protein S28	Trem-like transcript 1 protein	Platelet basic protein	Protein S100-A12
11	Neutrophil gelatinase-associated lipocalin	Prenylcysteine oxidase-like	Hemoglobin subunit alpha	Apolipoprotein A-I
12	Cathelicidin antimicrobial peptide	Non-histone chromosomal protein HMG-17	C-C motif chemokine 5	Cathepsin G

13		Neutrophil defensin 1	Neutrophil defensin 1
14		Bactericidal permeability-increasing protein	Band 3 anion transport protein

Legend: blue-greenish, hub-bottleneck proteins
Abbreviations: BD-D, bipolar patients in a depressive state; BD-M, bipolar patients in a manic state; MDD, major depressive disorder; SZ, schizophrenia

Table 9 Characteristics of the clusters and results of the functional enrichment analyses for biomarker candidates for differential diagnosis of MDD and BD-D patients.

Cluster	# Nodes	# Edges	# Genes	Category	Description	q-value
1	5	7	5	Reactome Pathways	Neutrophil degranulation	0,0055
			5	GO Component	Extracellular region	0,0086
			5	GO Component	Secretory granule lumen	0,007
			4	GO Component	Specific granule lumen	0,0086
			4	Reactome Pathways	Antimicrobial peptides	0,0099
2	4	6	4	UniProt Keywords	Chromosome	0,002
			4	UniProt Keywords	Hydroxylation	0,002
			4	UniProt Keywords	DNA-binding	0,0032
			4	GO Component	Nucleosome	0,0044
			4	UniProt Keywords	Methylation	0,0072
			3	UniProt Keywords	Nucleosome core	0,0072
			3	KEGG Pathways	Alcoholism	0,0081
			3	KEGG Pathways	Systemic lupus erythematosus	0,0081
			3	Pfam	Core histone H2A/H2B/H3/H4	0,0081
			3	Pfam	Histone-like transcription factor (CBF/NF-Y) and archaeal histone	0,0081
			3	UniProt Keywords	Citrullination	0,0094
			3	GO Component	Nuclear chromatin	0,0102
			4	UniProt Keywords	Nucleus	0,0159
			4	GO Component	Nuclear lumen	0,0192
			3	InterPro Domains	Histone H2A/H2B/H3	0,0203
			3	InterPro Domains	Histone-fold	0,0203
			4	GO Process	Chromatin organization	0,0206
			2	KEGG Pathways	Necroptosis	0,0243
			2	Pfam	C-terminus of histone H2A	0,0243
			4	GO Function	DNA binding	0,0373
			3	UniProt Keywords	Isopeptide bond	0,0412
			4	UniProt Keywords	Acetylation	0,0469
			3	UniProt Keywords	Ubl conjugation	0,0469
3	4	6	4	SMART Domains	Intermediate filament protein	1,3E-4
			4	Pfam	Intermediate filament protein	2,5E-4
			4	InterPro Domains	Intermediate filament protein	7,5E-4
			4	InterPro Domains	Intermediate filament protein, conserved site	7,5E-4
			4	InterPro Domains	Intermediate filament, rod domain	7,5E-4
			4	Reactome Pathways	Formation of the cornified envelope	7,5E-4
			4	UniProt Keywords	Coiled coil	0,0019
			4	UniProt Keywords	Intermediate filament	0,0019
			4	UniProt Keywords	Keratin	0,0019

			4	GO Component	intermediate filament	0,005
			3	UniProt Keywords	Ichthyosis	0,0076
			3	GO Component	Cornified envelope	0,009
			4	GO Function	Structural molecule activity	0,0098
			3	GO Function	Structural constituent of epidermis	0,0112
			4	UniProt Keywords	Disease mutation	0,0128
			4	GO Process	Cornification	0,0179
			2	Pfam	Keratin type II head	0,0182
			3	GO Process	Peptide cross-linking	0,0182
			2	InterPro Domains	Keratin, type I	0,0273
			2	InterPro Domains	Keratin, type II	0,0273
			2	InterPro Domains	Keratin type II head	0,0273
			2	KEGG Pathways	Estrogen signaling pathway	0,0364
			2	UniProt Keywords	Palmoplantar keratoderma	0,039
4	4	3	NA	NA	NA	NA
5	4	3	NA	NA	NA	NA
6	3	2	NA	NA	NA	NA
7	3	2	2	SMART Domains	Trypsin-like serine protease	0,0192

Abbreviations: BD-D, bipolar patients in a depressive state; MDD, patients with major depressive disorder; NA, not applicable

Table 10 Characteristics of the clusters and results of the functional enrichment analyses for biomarker candidates for differential diagnosis of BD-M and SZ patients.

Cluster	# Nodes	# Edges	# Genes	Category	Description	q-value
1	15	61	15	GO Process	Exocytosis	6,17E-5
			15	GO Process	Immune response	6,19E-5
			15	GO Process	Neutrophil activation	6,17E-5
			14	GO Process	Neutrophil degranulation	7,25E-5
			14	Reactome Pathways	Neutrophil degranulation	7,48E-5
			15	Reactome Pathways	Immune System	1,6E-4
			14	GO Component	Extracellular region	3,6E-4
			14	GO Component	Secretory granule	3,6E-4
			12	GO Component	Secretory granule lumen	3,6E-4
			14	GO Component	Cytoplasmic vesicle part	3,6E-4
			13	GO Component	Cytoplasmic vesicle lumen	3,6E-4
			12	GO Process	Defense response	0,0025
			11	GO Component	Extracellular space	0,0043
			14	GO Component	Intracellular organelle lumen	0,0043
			13	GO Process	Response to external stimulus	0,0072
			14	GO Process	Response to stress	0,0082
			12	UniProt Keywords	Secreted	0,0165
			12	GO Process	Multi-organism process	0,0177
			11	GO Process	Response to other organism	0,0187
			9	UniProt Keywords	Antimicrobial	0,0201
			11	UniProt Keywords	Disulfide bond	0,0201
			10	GO Process	Response to bacterium	0,0236
			9	GO Process	Defense response to bacterium	0,0272
			8	UniProt Keywords	Antibiotic	0,0287

			6	GO Component	Tertiary granule lumen	0,0298
			9	GO Process	Antimicrobial humoral response	0,0424
2	9	15	4	Pfam	Globin	0,0235
			8	UniProt Keywords	Acetylation	0,0338
			6	GO Component	Cytosolic part	0,0376
3	3	3	3	Reactome Pathways	Interferon gamma signaling	0,0051
			2	SMART Domains	Immunoglobulin C-Type	0,0202
			2	Reactome Pathways	ER-Phagosome pathway	0,0269
			2	Reactome Pathways	Endosomal/Vacuolar pathway	0,0269
			2	Reactome Pathways	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	0,0269
			2	Reactome Pathways	Interferon alpha/beta signaling	0,0269
			2	Reactome Pathways	Antigen Presentation: Folding, assembly and peptide loading of class I MHC	0,0269
4	3	2	NA	NA	NA	NA
5	4	3	NA	NA	NA	NA
6	5	7	3	KEGG Pathways	Cholesterol metabolism	0,0139
			4	GO Component	High-density lipoprotein particle	0,0192
			5	GO Component	Extracellular region	0,0389
			3	GO Component	Very-low-density lipoprotein particle	0,0389
			3	GO Component	Spherical high-density lipoprotein particle	0,0389
			3	GO Component	Chylomicron	0,0389
7	3	3	NA	NA	NA	NA

Abbreviations: BD-M, bipolar patients in a manic state; NA, not applicable; SZ, schizophrenia patients

4. Discussion

As has already been recognised for decades, the lack of objective diagnostic biomarkers for major psychiatric disorders impedes psychiatric healthcare. Therefore, many studies aimed to fill this gap by assessing various biological features of psychiatric patients. However, psychiatric diagnostics remains hampered by the high degree of subjectivity involved in the process, which may partially be a result of the *modus operandi* implemented in previous studies. To this day, biomarker discovery was mainly based on differentiating patients with a defined psychiatric disorder from healthy controls rather than differentiating psychiatric patients with different, though clinically similar, disorders. The latter approach will generate biomarkers that are clinically useful because the challenge within psychiatric diagnostics lies within differentiating these patients from each other. Hence, this thesis was focussed on biomarker discovery for differential diagnosis of, on the one hand, MDD and BD-D patients and, on the other hand, BD-M and SZ patients.

An optimised protocol is necessary to generate reliable results and increase the likelihood of finding biomarker candidates. Therefore, the first objective of this thesis was to optimise the protocol for sample preparation of plasma samples as well the LCMS protocol. However, one should keep in mind that the optimisation phase of this study was rather limited and thus does not provide a solid optimised protocol for sample preparation of plasma. Subsequently, the performance of the optimised protocol for plasma samples was compared to the standard protocol for PBMCs to assess which sample type provides a better sample source in terms of biomarker discovery.

Additionally, this thesis included the analysis of protein profiles obtained from PBMCs of the previously mentioned patient populations. The approach used in this thesis allowed to identify pathology-related biomarker candidates by overlapping $\log_2$ fold change expression ratios between patient groups with those of patient groups compared to healthy controls. Moreover, pathway and network analysis enabled to extract biological information from these biomarker candidates.

4.1 Performance of the ProteoSpin™ and Pierce™ Depletion Column

The success of biomarker discovery partially relies on the number of biomarker candidates that can be detected. The more proteins are identified, the greater the change that one of the identified proteins is differentially expressed and thus a potential biomarker candidate. Therefore, the number of identifications is an important measure to consider when optimising a protocol for this purpose. Moreover, a high quality biomarker should be detected in most (if not all) samples in a stable manner. Hence, the number of proteins that were detected in all runs and the CV at multiple levels were also assessed.

To the best of my knowledge, this thesis was the first to compare the performance of the ProteoSpin™ and Pierce™ column. The ProteoSpin™ column outperformed the Pierce™ column at all levels. Samples that were prepped with the ProteoSpin™ column generated a larger number of identifications after three runs, a larger number of proteins identified in all runs of a sample and a larger number of proteins with a CV below 20%. Proteins detected with a CV above 20% would not be reliable biomarker candidates and would thus not be withheld in the discovery phase of a study aiming to identify biomarkers.

To get an estimate of the variation on the inter-run variability across different samples, the range of the inter-run variability was assessed at protein level by subtracting the highest inter-run CV with the lowest. Conflicting with the other results, the median range of the inter-run variability of samples

depleted with the Pierce™ column was lower compared the one of samples depleted with the ProteoSpin™ column. As can be deduced from Figure 9, this may be caused by one sample depleted with the ProteoSpin™ column showing a deviant distribution of the inter-run variability compared to the other samples that were depleted with this column. The majority of proteins detected in this sample had a CV above 60% whereas the majority of proteins were detected with an inter-run variability below 40% in the other samples depleted with the ProteoSpin™ column. When this sample was removed from the dataset, the median range of the inter-run variability dropped below 20%. Hence, the deviating distribution of the inter-run variability in one sample caused the large median range of the inter-run variability in samples depleted with the ProteoSpin™ column. A larger sample size is necessary to conclude whether the distribution of the inter-run variability in this sample is an exception or a common phenomenon.

When data were analysed with the MaxLFQ algorithm, the performance of both depletion columns was similar. The effect of MaxLFQ was assessed because it showed to quantify protein abundances more accurately [168], as evidenced by the decreased median inter-run variability across all samples and range of the inter-run variability for both depletion columns. However, it also decreased the number of identifications after three runs and the number of proteins that were detected in all runs for both depletion columns. Nevertheless, the MaxLFQ algorithm decreased the overall variation at all levels and thus generated more reliable results. This is consistent with results obtained by Zhao et al., who analysed seven commonly-used label-free quantification methods. They reported that MaxLFQ quantified proteins more accurately and precisely than the other methods [199]. The same results were obtained in various other studies [168, 200-202]. However, the authors of a recently developed (April 2020) method reported that their method outperforms MaxLFQ when data is used for differential expression analysis, as will be the case in a biomarker discovery phase [203]. Thus, this method should be investigated in a benchmarking study that is not performed by any of the authors of included methods to decide which method performs best.

Although COVID-19 interfered with the data collection that was necessary to conclude which protocol yields better results for biomarker discovery, the ProteoSpin™ column and the MaxLFQ algorithm seem to be the better choice based on the results that are available at the moment. As the comparison of these depletion columns was never made before, additional experiments are necessary to support this observation. These experiments should at least include samples from different test subjects. This way, the inter-individual variability can also be assessed. Additionally, the inter-prep variability should be assessed in a more profound manner by using multiple aliquots of different test subjects. Using a larger sample size will also allow to integrate the results of both depletion columns and analytical algorithms via a two-way ANOVA. The small sample size available in this thesis is a major limitation which caused that certain assumptions made by this test (e.g. normality) could not be guaranteed. Therefore, differences were assessed via a non-parametric test.

4.2 Effect of the Length of the LC Gradient on the Number of Identifications

As stated earlier, the number of identifications is an important measure to consider when optimising a protocol for biomarker discovery. Even though the speed and sensitivity of mass spectrometers improved in the last years, peptides in complex samples remain undetected as many peptides simultaneously elute from the LC column [204]. Therefore, several chromatographic parameters have been varied and their effect on the number of identifications has been investigated intensively in the last years [164, 205, 206]. Increasing the length of the gradient that is applied to elute the peptides proved to positively influence the number of identifications [164, 165]. However, this modification of the protocol also increases the total running time, which could become very large when many samples have to be analysed. Therefore, a trade-off should be made between the gradient length and the total running time. As total running time depends on the number of samples, the number of runs per sample

and the gradient length, it is in essence a trade-off between the coverage of the true biological heterogeneity, the quantitative precision and the number of identifications.

The positive effect of increasing the length of the LC gradient on the number of identifications can be explained by the effect of the length of the LC gradient on the peak capacity, which is correlated with the number of identifications. The peak capacity of an LC column is defined by the number of components that can be separated and dependent on the length of the LC gradient and the column length [206, 207]. Both Kocher [206] and Gilar [207] et al. found that the peak capacity increases with an increasing gradient length. Moreover, Kocher et al. identified a linear relation between the peak capacity and the number of identifications and thus also an increased number of identifications with an increasing gradient [206].

This part of the experimental set-up was largely affected by the COVID-19 lockdown. Therefore, the results obtained during this thesis are not sufficient to draw any conclusions, even though there seems to be an increasing trend of the number of identifications dependent on the gradient length reaching a maximum at 240 minutes. An unexpected phenomenon occurred at 180 minutes, where the number of identifications dropped to 160 compared to 186 when a gradient of 155 minutes was applied. This may potentially be caused by the fact that samples used to test the 180-440 minutes gradients underwent 2 freeze-thaw cycles whereas the sample used to test the 155 minutes gradient only had 1. However, multiple studies have investigated the effect of freeze-thaw cycles on protein abundances and reported that these were not affected after up to eight cycles. Unfortunately, the number of identifications dependent on the number of freeze-thaw cycles was not reported in these studies [208, 209]. Nevertheless, this phenomenon may also be explained by the fact that the number of identifications varies in between runs, as evidenced by the variation on the number of identifications when samples that were used in this experiment were analysed with the gradient of 155 minutes (protocol described in section 2.1.1.4). These samples yielded between 180 and 189 identifications per run. Based on the results available at the moment (notwithstanding the fact that these are not sufficient!), a gradient of 240 minutes would be the best choice considering that a longer gradient does not increase the number of identifications further. However, this decision should also take into account whether or not upscaling the length of the LC gradient is feasible in terms of increased running time and costs associated with it, which is dependent on the sample size that will be used in the biomarker discovery phase.

The following steps were part of the original experimental set-up but could not be executed due to COVID-19. Hence, they should be taken to further optimise this part of the protocol. First, the effect of the gradient length on the number of identifications should be assessed using samples from multiple test subjects, multiple aliquots per test subject and multiple runs per aliquot, offering the possibility to objectively quantify variation at multiple levels. This approach allows to decide which gradient length is best in terms of the number of identifications, variation on the number of identifications and total running time. Moreover, this information should be combined with the estimate of variation on protein intensities, both on a technical and biological level. This way, the number of required runs and aliquots per test subject that are necessary to allow accurate quantification of protein abundances can be determined. Furthermore, this enables the possibility to perform a power analysis, providing information about the sample size that is essential to detect a significant difference. Second, the optimisation phase should mimic the situation of a biomarker discovery phase as much as possible. In the biomarker discovery phase, many samples will have to be prepped before they can be analysed by the LCMS. It is not feasible to prep all samples simultaneously and immediately load them onto the LCMS. Therefore, prepped samples will have to be stored at -20°C and thawed before being loaded

into the LCMS. Although this thesis already implemented this approach, the variation on the number of freeze-thaw cycles was inevitable due to time limitations, which may have affected the results. As all internships within university buildings were suspended because of the COVID-19 outbreak, a choice had to be made between not assessing this parameter at all or using samples with a different number of freeze-thaw cycles.

4.3 Should we opt for Plasma or PBMCs as a Source for Biomarker Discovery?

As biomarkers should ideally be easily accessible, many studies focused on biomarker discovery using plasma. However, proteomic analysis of plasma is associated with some challenges because of its high complexity and large dynamic range of protein abundances [157, 158]. PBMCs might be an interesting sample source for biomarker discovery because they yield a higher number of proteins than plasma [121]. Additionally, they are proposed to be a neural probe, which makes them specifically interesting for biomarker discovery in brain disorders, including psychiatric disorders [127]. Therefore, the results obtained from LCMS analysis of plasma and PBMCs were compared with each other in terms of quantitative and qualitative endpoints.

Similar to the results obtained by Končarević and colleagues [121], LCMS analysis of PBMCs resulted in a higher number of identifications compared to plasma. However, the number of proteins identified in this thesis is substantially lower for both sample sources. Končarević and Coppens and colleagues identified more than 4000 proteins in PBMCs whereas more than 1000 proteins are regularly identified in plasma [121, 147, 210-212]. This discrepancy in the number of identifications can be explained by the fact that these studies use extensive fractionation systems. Although these approaches yield a higher number of identifications compared to the approach used in this thesis, these approaches also substantially increase running time, which is not feasible when a large sample size should be analysed (as will be the case in a study focused on biomarker discovery). Furthermore, an increased running time inevitably limits the number of runs that can be applied for each sample. As this is an important step to reach high quantitative precision, applying multiple runs per sample is as important as the number of identifications. Studies using a one-dimensional fractionation approach reported less protein identifications (150-500 for plasma, 400-2000 for PBMCs) [213-217], which were comparable with the 348 and 1077 proteins that were identified in plasma and PBMCs this thesis.

Although the quantification of protein abundances was characterised by a larger overall variation in PBMCs than in plasma, the number of proteins that are useful for biomarker discovery (i.e. proteins that are quantified with an inter-run variability below 20%) was larger in PBMCs. Similar to plasma, this number increased when data were analysed using the MaxLFQ algorithm. Therefore, PBMCs and the MaxLFQ algorithm seem to be the better choice for biomarker discovery even though the data collected during this thesis are not sufficient to draw solid conclusions.

Specifically, two important measures could not be compared due to the limited availability of data, namely the inter-individual and intra-individual variability. Although these are intellectual measures and not technical ones (like inter-run and inter-prep variability), they are valuable when deciding which sample source provides the best characteristics for biomarker discovery. As previously mentioned, a biomarker should allow for a stable detection to be clinically useful and thus, variation on protein abundances across individuals and within the same individual over time should be limited. Proteins showing a highly variable abundance across individuals and within the same individual over time would impede the detection of true biological differences and should thus be excluded for biomarker discovery [218]. As two test subjects were used to test the standard protocol for sample preparation

of PBMCs, the inter-individual variability could be assessed in this sample source, albeit in a very superficial manner. The mean inter-individual variability of protein abundances quantified in this thesis was 48,83% when data were analysed with the standard MaxQuant algorithm and dropped to 24,58% when the MaxLFQ algorithm was used. The latter was comparable with results obtained by Maes et al. (28%), who investigated the inter-individual variability on protein abundances in PBMCs of a control elderly population consisting of 24 individuals [218]. However, the results of Maes et al. were obtained with 2D-gel electrophoresis (2D-GE) and might thus not be directly comparable with results obtained via another technique. Nevertheless, the strong decreased mean inter-individual variability when data were analysed with the MaxLFQ algorithm is another argument for the use of this algorithm instead of the standard MaxQuant algorithm. The inter-individual variability of plasma proteins has also been estimated before via 2D-GE using plasma samples derived from 11 individuals. The researchers reported a median inter-individual variability of protein abundances in plasma of 23% [219].

4.4 Interesting Biomarker Candidates

This thesis also included data analysis of protein expression levels of patients with major psychiatric disorders that were obtained in a proof-of-concept study. The aim of this proof-of-concept study was to assess the potential of PBMCs to serve as a sample source for biomarker discovery in major psychiatric disorders. The sample size is too small to validate identified biomarker candidates because it does not represent the heterogeneity in true patient populations. Most likely, biomarker candidates will change when this study is replicated with a decent sample size. However, data were analysed as if it was obtained from a larger sample size to acquire hands-on experience with the bioinformatic methods that would have been used in the original experimental set-up. Thus, the results of this study should be taken with a (large) grain of salt. Nevertheless, the results provide valuable information regarding the usability of PBMCs for biomarker discovery in psychiatry and may pave the way towards a shift of which sample source to use for this purpose.

The most interesting biomarker candidates for differential diagnosis are those that show to be contra-regulated in the disorders they should discriminate. Unfortunately, this was only the case for two proteins that showed to be differentially expressed between BD-M and SZ patients. In contrast, 16 biomarker candidates for differential diagnosis of MDD and BD-D patients and 21 biomarker candidates for differential diagnosis of BD-M and SZ patients were differentially expressed in the same direction for both experimental groups. In this case, protein expression levels should be sufficiently distinct between the experimental groups to allow for their discrimination, as was a criteria to be defined a biomarker candidate in this study. Although proteins that do not meet this criteria cannot distinguish the experimental groups, they may provide valuable insight in overlapping underlying disease mechanisms. 69 of such proteins were differentially expressed in MDD and BD-D patients while this was the case for 60 proteins when comparing BD-M and SZ patients. 41 of these proteins were differentially expressed in all patient groups and could thus be potentially a result of underlying disease mechanisms that major psychiatric disorders have in common.

4.4.1 Contra-Regulated Proteins

As previously mentioned, two biomarker candidates were found to be contra-regulated, both for differential diagnosis of BD-M and SZ patients. These biomarker candidates are the most interesting because they indicate that a biological process linked to that protein might be differentially affected in the disorders of interest. Interestingly, the identified contra-regulated proteins were among the biomarker candidates with the highest discriminatory potentials, as expected. HLA-B7 was ranked number three based on the magnitude of the $\log_2$ fold change expression ratio between BD-M and SZ

patients whereas APOC3 was ranked sixth. Even though literature does not provide a direct link between HLA-B7 and psychiatric disorders, it can be linked to the immune hypothesis which postulates that psychiatric disorders may be caused by immunological deviations (for review see [220, 221]). This matter is discussed more thoroughly in section 4.5.

APOC3 is a protein that binds various types of lipids, such as chylomicrons and very low density lipoproteins, and is in this way a key player in the transport and metabolism of these hydrophobic molecules [222]. Whereas elevated plasma levels of this protein have already been linked to an increased risk of cardiovascular disease [223-225], more recent studies also found altered levels in psychiatric disorders. The present study identified significantly increased expression levels of APOC3 in SZ patients. These results are consistent with those of Boiko et al., who also reported a correlation between APOC3 serum levels and the presence of metabolic syndrome in SZ patients [226], and Domenici et al., who also detected altered levels in MDD patients [114]. Additionally, Knöchel and colleagues stated that this increase was also present in plasma of the same patient population as well as BD patients [227]. However, this study detected decreased APOC3 levels in BD-M patients, as was also reported by Herberth et al. in general BD patients [146]. APOC3 levels remained unaltered in MDD and BD-D patients.

4.4.2 Hub-Bottleneck Proteins in the Protein-Protein Interaction Networks of Biomarker Candidates

Hub-bottleneck proteins have a high degree as well as a high betweenness centrality and are thus likely to be essential elements of the network [192, 193]. Therefore, these biomarker candidates may represent critical communication points between affected biological processes. According to the literature, hub-bottleneck proteins in the protein-protein interaction network of biomarker candidates for differential diagnosis of MDD and BD-D patients have not been directly linked to psychiatric disorders yet and are thus not discussed below. Likewise, hub-bottleneck proteins in the protein-protein interaction network of biomarker candidates for differential diagnosis of BD-M and SZ patients that did not show any link to psychiatry are not discussed[8]. Interestingly, all hub-bottleneck proteins that have already been linked to psychiatric disorders are involved in the immune system.

Haptoglobin, which among others binds free haemoglobin and acts as a positive acute phase protein [228], was ranked the number one hub and number two bottleneck protein within the protein-protein interaction network of biomarker candidates for differential diagnosis of BD-M and SZ patients. Consistent with the results obtained in this thesis, numerous studies reported increased haptoglobin levels in plasma as well as serum and whole blood of SZ patients [107, 229-234]. However, this thesis did not identify aberrant abundances of this protein in MDD, BD-D and BD-M patients whereas other studies reported increased levels in plasma of these patients [114, 120, 232, 235, 236].

Two other hub-bottleneck proteins within the network of biomarker candidates for differential diagnosis of BD-M and SZ were cathelicidin antimicrobial peptide (CAMP) and neutrophil gelatinase-associated lipocalin (LCN2), both involved in the innate immune response against bacterial infection [237, 238]. These proteins were up-regulated in all patient groups compared to controls, but most in SZ patients where they demonstrated the highest and second highest $\log_2$ fold change of all detected proteins. Interesting is that these proteins were able to differentiate MDD and BD-D patients as well

[8] Proteins were only considered to be linked to psychiatry when experimental evidence was provided for that specific protein. Evidence linking family members of the protein to psychiatry was not considered a direct link due to a restricted number of pages.

as BD-M and SZ patients from each other and that they showed a consistent rank order of the magnitude of the $\log_2$ fold change. The protein abundances of these biomarker candidates differed the most in SZ followed by MDD, BD-M and BD-D. This may offer the possibility to differentiate these four patient groups by a single biomarker. However, this was not examined in this thesis because it went beyond the scope of this project. Various studies have already found deviant abundances of both proteins in psychiatric patients. Similar to our results, the majority of these studies report an increase of protein abundances in MDD, BD and SZ [239-242]. However, some studies also demonstrated decreased CAMP levels and a large variation on this measure in patient as well as control groups [239, 240, 243]. Therefore, CAMP might not be a reliable biomarker candidate.

Another hub-bottleneck protein that has already been linked to psychiatric disorders is matrix metalloproteinase-9 (MMP9), which is mainly involved in the structural organisation of extracellular matrix but can also be linked to the immune system [244-246]. This protein was found to be up-regulated in all patient groups compared to controls, but only showed differential expression between BD-M and SZ patients and not between MDD and BD-D patients. The involvement of MMP-9 in psychiatric disorders has been extensively investigated, both on gene and protein expression level. All studies reported results consistent with those obtained in this thesis [112, 114, 247, 248]. Among others, two studies reported an up-regulation of MMP9 gene expression in PBMCs of SZ patients [249, 250]. These increases might be caused by the presence of a polymorphism in the MMP9 gene. Researchers demonstrated that the polymorphism with a higher transcriptional activity (T allele) was more frequently detected in bipolar patients compared to controls, in which the other polymorphism (C allele) was more common [248]. However, the opposite is true for SZ patients while they also exhibit increased gene and protein expression of MMP9 [114, 249-251].

Protein S100A12 (S100A12), a calcium-binding protein mainly involved in the immune system [252], and neutrophil defensin 1 (DEFA1), released during the immune response against pathogens [253], were both identified as biomarker candidates for differential diagnosis of BD-M and SZ patients and hub-bottleneck proteins. Additionally, both protein were up-regulated in all patient groups compared to controls. Multiple studies have also detected increased levels of this protein or its gene expression in MDD, BD and SZ patients [254-257], including two studies investigating these measures in PBMCs of SZ patients [152, 258].

Limited evidence was found to link C-C chemokine 5 (CCL5) to psychiatric disorders. Although this study only detected altered expression levels of CCL5 in BD-M patients, a review did not identify any links between CCL5 and BD nor did the literature search performed by the author of this thesis [259]. Other studies detected CCL5 abundances in MDD and SZ patients that deviate from healthy controls. However, some studies reported increased expression values whereas others report a decrease [107, 114, 260-262].

4.5 Affected Pathways, Processes and Systems

The results of this thesis clearly link altered immune functioning to psychiatric disorders. First, multiple proteins that were identified as biomarker candidates for differential diagnosis (e.g. HLA family, DEF1A, S100A12, CAMP and LCN2) play an important role the immune system. Second, pathway analysis of expression values of MDD and BD-D patients implicated that multiple immune- related gene sets were significantly enriched and thus differentially altered between these patient groups. Finally, clustering and subsequent functional enrichment analysis of both protein-protein interaction networks of biomarker candidates for differential diagnosis identified multiple clusters within those networks that

are involved in the immune system. Such alterations of immune-related processes in psychiatric disorders have been recognised for more than a century. The earliest studies date from the 19th century and reported there may be a relation between psychotic and acute infectious illnesses [220]. Nowadays, the 'immune hypothesis' is one of the many theories put forward aiming to explain the aetiology of major psychiatric disorders (for review see [221]). Nevertheless, the results of this thesis were obtained from cells that are key players of the immune system and should thus be interpreted carefully when linking them to the 'immune hypothesis'.

Next, pathway analysis of expression values of MDD and BD-D patients revealed that various translation- and post-translational modification (PTM)-related gene sets were enriched. In the top 10 of significantly enriched gene sets, 7 are related to these biological functions. Moreover, they are all upregulated, implicating that expression levels of proteins involved in these processes are higher in MDD compared to BD-D. Various studies further support the fact that translation as well as post-translational modification might be affected in MDD patients [263-268]. Nevertheless, functional enrichment analysis of clusters within the network of biomarker candidates for differential diagnosis of MDD and BD-D patients did not yield the same results. As pathway analysis takes into account expression values of all detected proteins while network analysis only works with biomarker candidates, this discrepancy may potentially be explained by the fact that individual proteins involved in enriched gene sets related to translation and PTM might not have a large effect whereas their combined effect can be more profound.

Another biological function that was represented by many significantly enriched gene sets when expression values of MDD and BD-D patients were compared was the interaction between cells. Gene sets included in this overarching term include cell adhesion in terms of the organisation of extracellular matrix as well as in terms of platelet aggregation. Whereas platelet aggregation could not be linked to psychiatric disorders and may be a result of contamination, cell adhesion-related gene sets have already been identified to be significantly enriched in MDD patients [269, 270]. However, another study stated that expression levels of cell-adhesion molecules were altered in PBMCs of BD-D patients but not in MDD patients compared to controls [271]. In the present study, both patient groups demonstrated differential expression profiles leading to the identification of cell adhesion-related gene sets. In the brain, such gene sets are involved in neural plasticity [272, 273], which has also been proposed to be a pathophysiological mechanism underlying psychiatric disorders [274-276].

Surprisingly, pathway analysis of expression values in MDD and BD-D patients identified several significantly enriched gene sets that are related to the muscular system, more specifically gene sets related to cardiomyopathy and muscle morphogenesis. Although these gene sets were first considered to be rather unexpected results, a literature search revealed that similar gene sets have already been identified as significantly enriched in MDD patients and people with suicidal behaviour [267, 269, 270]. Moreover, researchers found that one particular single-nucleotide polymorphism (SNP) located within the cardiomyopathy associated 5 (CMYA5) gene was associated with a risk for MDD and SZ [277-279]. However, the results in this thesis implicate that the dysregulation of the muscular system lies within BD-D and SZ patients compared to controls whereas these gene sets were not detected to be significantly enriched in MDD and BD-M patients compared to controls. Nevertheless, people suffering from these major psychiatric disorder are at increased risk to develop cardiovascular diseases [16, 280-286].

Besides pathway analysis, protein profiles were also analysed via network analysis. In the network of biomarker candidates for differential diagnosis of MDD and BD-D patients, several biomarker

candidates, of which four are core components of histones[9], clustered together and generated significantly enriched terms related to epigenetics after functional enrichment analysis. These terms include methylation and acetylation, which are the most well-known epigenetic processes and can both occur on histones. Epigenetics-related terms as well as expression levels of members of the histone family have been reported to be altered in major psychiatric disorders [287-290]. These alterations could be related to each other because epigenetics influences expression levels without changing the DNA sequence itself. Interesting is that epigenetics can modify gene and thus also protein expression in response to environmental stimuli, which may be an explanation for the episodic nature of many psychiatric disorders [291].

Functional enrichment analysis of cluster 6 in the protein-protein interaction network of biomarker candidates for differential diagnosis of BD-M and SZ patients highlighted that there were some differences in these patient groups with regard to lipid metabolism. Indeed, literature suggests that lipid metabolism is dysregulated in SZ as well as BD patients [292-296]. In SZ patients, an altered lipid metabolism might be a result of their antipsychotics used to alleviate psychotic symptoms as the development of metabolic syndrome is a major side effect of most antipsychotics [297]. However, an altered lipid metabolism was also detected in drug-naïve SZ patients [295, 296]. The cluster generating lipid metabolism-associated terms consisted of three members of the apolipoprotein, which have been reported to be differentially regulated in psychiatric patients many times [108, 109, 120, 226, 227, 298]. However, these studies reported opposing results with regard to whether they are up- or down-regulated.

Lastly, four proteins of the keratin family were identified as biomarker candidates for differential diagnosis of MDD and BD-D patients and clustered together in the protein-protein interaction network thereof. However, no evidence was found that members of the keratin family are expressed in PBMCs. Hence, the detection of these proteins is probably a result of contamination via skin cells [299].

4.6 Protein-Protein Interactions between Biomarker Candidates for Differential Diagnosis

Protein-protein interactions were more profound for biomarker candidates for differential diagnosis of BD-M and SZ patients than those for differential diagnosis of MDD and BD-D patients. In the former network, the majority of biomarker candidates were connected with each other (Figure 19), either directly or indirectly[10], whereas biomarker candidates are only connected to few others in the latter network (Figure 18). This is further supported by the results of NetworkAnalyzer, which indicate that the number of edges, highest degree and average number of neighbours are higher in the network of biomarker candidates for differential diagnosis of BD-M and SZ patients. The protein-protein interaction network obtained from these biomarker candidates is consistent with the theory of the so called 'disease module hypothesis'. This hypothesis states that disease genes or gene products tend to interact and cluster with each other instead of showing no interaction at all [300]. Moreover, the 'local hypothesis' postulates that directly interacting proteins may be involved in the same biochemical process [301] and Barabási et al. hypothesised that disorders are mostly caused by the interaction of affected pathobiological processes [192].

[9] Histones are core components of nucleosomes, in which DNA is packed around histones to reduce its length.
[10] A direct interaction is present between neighbours while proteins that interact indirectly are connected to each other via (a) node(s) in between.

This raises the question if there could be a major network of underlying biological differences between BD-M and SZ patients while this is not the case for MDD and BD-D patients. However, approximately 75% of all possible protein-protein interactions are not detected or examined yet [302]. Hence, interactions between the smaller protein-protein interaction networks created by biomarker candidates for differential diagnosis of MDD and BD-D patients may not have been identified yet. This hypothesis is not in line with the results obtained via pathway analysis. As previously mentioned, network analysis demonstrated that the majority of biomarker candidates for differential diagnosis of BD-M and SZ patients interact with one another and thus indicated that these proteins might be involved in the same biological processes, according to the 'local hypothesis'. It was therefore suspected that pathway analysis would identify significantly enriched gene sets that were similar to those identified with the STRING enrichment of clusters in the protein-protein interaction network. However, pathway analysis did not yield any significantly enriched gene sets when expression values of BD-M and SZ patients were compared. The opposite was true when expression values of MDD and BD-D patients were compared where pathway analysis identified 499 significantly enriched gene sets.

4.7 Future Perspectives

Ideally, a biomarker discovery study should include the follow-up of drug-naïve patients[11]. This study design would allow for the protein profiling of samples of the same patients before and after treatment and therefore creates the possibility to only withhold biomarker candidates when they are differentially expressed in both conditions. Moreover, it would enable to evaluate the stability of biomarker candidates throughout time. This approach guarantees that the biomarker (panel) can discriminate patients independent of their drug status or the time point in the disease course at which the diagnostic test is performed, which may be specifically important when the biomarker (panel) has only just been implemented in the clinic due to the fact that many patients probably won't be drug-naïve at the time point they are biologically diagnosed. Additionally, future studies should, preferably, simultaneously investigate protein profiles of the same experimental groups as were included in this thesis when aiming to identify biomarker candidates for differential diagnosis of MDD and BD-D patients, on one hand, and/or BD-M and SZ patients, on the other. This approach allows that biomarker candidates are selected based on their potential for differential diagnosis between clinically similar pathologies, which is the most challenging in clinical practice.

In the present study, biomarker candidates were defined as proteins that showed to be differentially expressed between two patient groups as well as between either of those patient groups and HCs. Therefore, this approach enables that clinical biomarking can be performed following two strategies, as stated by Coppens and colleagues [147]. A first strategy (Strategy A) requires that the protein abundances of biomarker candidates are determined in a large cohort of HCs, allowing a reference range and threshold value to be set for each biomarker candidate. This strategy can be used for proteins which show to be contra-regulated or differentially regulated in either of the pathologies compared to HCs. In the latter case, the fold change expression ratio should be less than 1 SD to the mean fold change expression ratio in one of the patient groups and more than 2 SD in the other (i.e. considered differentially expressed in this thesis). In each list of biomarker candidates for differential diagnosis identified in this thesis, 35 proteins meet this criteria. Additionally, two biomarker candidates for differential diagnosis of BD-M and SZ patients were contra-regulated, meaning that 72 biomarker

[11] A drug-naive patient is considered as a patient that has not received drugs prescribed to alleviate symptoms related to the disorders of interest yet.

candidates identified in this thesis would be eligible for further validation using the strategy based on a HC reference range.

A second strategy (Strategy B) for biomarker selection and validation would be applicable to proteins that show to be significantly dysregulated in both pathologies compared to controls as well as in comparison with each other. Similar to the first strategy, this strategy also requires that reference ranges of protein abundances are determined. However, instead of determining a reference range of protein abundances in healthy controls, this type of biomarker candidates demand that the reference ranges of protein abundances in both patient groups are set using a large cohort of patients. Biomarker candidates showing the greatest discriminatory potential between the two comparative pathologies would be best suited for this strategy.

Biomarker candidates that show to be significantly contra-regulated (validated via Strategy A) or significantly dysregulated in both pathologies (validated via Strategy B) take into account fold change expression ratios of both pathologies. Therefore, these types of biomarkers would probably prove to be more specific and sensitive than biomarker candidates that are only significantly dysregulated in one of the pathologies. Nevertheless, extensive follow-up research is necessary, no matter which type of biomarker candidate is selected for further validation. A first step consists of the protein profiling of PBMCs derived from a larger cohort to identify true biomarker candidates. Next, these biomarker candidates should be validated, preferably by using single or multiplex enzyme-linked immuno-assays as such techniques allow low-threshold implementation in the clinic. The following step consists of establishing reference ranges for biomarker candidates that passed through the previous steps. Depending on to which of the previously described strategies the biomarker candidate belongs, these reference ranges should be determined in either a large cohort of HCs (Strategy A) or both patient populations (Strategy B). In a final step, these reference ranges can then be used to assess the discriminatory performance of the biomarker candidate via the comparison of the biological and the 'subjective' diagnosis in a large cohort of patients. The subjective diagnosis should preferably be based on the opinion of at least two experienced psychiatrists[12] and the results of multiple diagnostic instruments and questionnaires. This way, the biomarker candidate's accuracy, sensitivity and specificity can be determined.

The abovementioned strategies and validation steps are only applicable for single biomarkers. However, single biomarkers have not been implemented in the clinic because they showed to be non-specific, possibly because of their inability to represent complex underlying disease mechanisms [118]. In case single biomarkers do not provide an accuracy, sensitivity and specificity sufficient for implementation in clinical practice, protein profiles of patients and HCs may be used to extract biomarker panels from, an approach that has already been applied successfully in various research areas [303-307]. If this approach still yields disappointing results, a last resort might be to start from the biological information instead of the diagnostic labels as we know them today, meaning that patients will be clustered together based on their protein profiles. These protein profiles might then be used to classify patients even though these classifications might not overlap with the current diagnostic categories. This approach has also been widely used, specifically in the context of disease subtyping, and yielded valuable information in many cases [308-311]. Both approaches make use of machine learning algorithms, supervised in the former case (e.g. support vector machine) and unsupervised in the latter (e.g. cluster analysis).

[12] A third psychiatrist should be consulted in case the opinion of only two psychiatrists is originally included and they do not agree on the diagnosis.

5. Conclusion

Major psychiatric disorders are among the most debilitating disorders, costing patients and global economy trillions of dollars each year. Although biomarker discovery in psychiatry has been a subject undergoing intense study in the last decades, biological measures are still lacking in psychiatric diagnostics. Therefore, the study designs that are generally used in biomarker discovery may have to be revised. Only few studies focused on differential diagnosis of major psychiatric disorders instead of comparing psychiatric patients only to healthy controls while the former is of utmost importance in clinical practice.

This thesis provided preliminary data indicating that breaking the golden standards in biomarker discovery may be considered. First, one should look beyond samples sources that are classically used in biomarker discovery. This thesis implicates that PBMCs may offer better characteristics for biomarker discovery than plasma. Furthermore, PBMCs generated a considerable number of proteins that are differentially expressed between psychiatric disorders with similar clinical presentations. Hence, they may prove to be a valuable sample source in biomarker discovery for differential diagnosis of major psychiatric disorders.

Extensive research is necessary to improve psychiatric healthcare, starting with biomarker discovery for diagnostic as well as prognostic and therapeutic purposes. Research should primarily be focused on solving the most challenging aspect of psychiatric healthcare, namely differential diagnosis of psychiatric disorders with clinically similar presentations. Next, these results can aid in the development of new therapeutics. By implementing objective, biological measures, the quality of psychiatric healthcare will increase, which will in turn lead to a decreased disease and economic burden caused by these debilitating afflictions.

6. Acknowledgements

This master's thesis was performed at the Scientific Initiative for Neuropsychiatric and Psychopharmacological Studies, which is a collaboration between the University of Antwerp and the University Psychiatric Centre Duffel, and the Receptor Biology Lab. I would like to thank my promotors, Prof. dr. Violette Coppens and Prof. dr. Manuel Morrens, for giving me the opportunity to explore my interests in psychiatric disorders and co-readers, Prof. dr. Stuart Maudsley and Prof. dr. Nathalie Cools, for their honest opinions and questions. I also sincerely thank Jobbe Goossens. Thank you for your advice and guidance throughout these seven months, thank you for telling me to relax from time to time and thank you for being such a great mentor! Hanne Leysen is also sincerely thanked for her guidance and expertise in the laboratory. All members of the Scientific Initiative for Neuropsychiatric and Psychopharmacological Studies and Receptor Biology Lab are also thanked for their hospitality. Other persons that deserve a big thank you are my parents, family and friends. Lastly, I would also like to thank Joris for his much needed emotional, and sometimes even technical, support.

7. References

1. *Global, regional, and national age-sex-specific mortality and life expectancy, 1950-2017: a systematic analysis for the Global Burden of Disease Study 2017.* Lancet, 2018. **392**(10159): p. 1684-1735.
2. Kay, J. and A. Tasman, *Mood disorders: Bipolar (Manic-Depressive) Disorders*, in *Essentials of Psychiatry*. 2006, John Wiley & Sons, Ltd. p. 556-572.
3. Kay, J. and A. Tasman, *Mood disorders: Depression*, in *Essentials of Psychiatry*. 2006, John Wiley & Sons, Ltd. p. 533-555.
4. Kay, J. and A. Tasman, *Schizophrenia and Other Psychoses*, in *Essentials of Psychiatry*. 2006, John Wiley & Sons, Ltd. p. 495-532.
5. Allsopp, K., et al., *Heterogeneity in psychiatric diagnostic classification.* Psychiatry Res, 2019. **279**: p. 15-22.
6. Health, W.M. *Prevention of Suicidal Behaviours: A Task for All*. Available from: http://www.who.int/mental_health/prevention/suicide/background
7. Cavanagh, J.T., et al., *Psychological autopsy studies of suicide: a systematic review.* Psychol Med, 2003. **33**(3): p. 395-405.
8. Chesney, E., G.M. Goodwin, and S. Fazel, *Risks of all-cause and suicide mortality in mental disorders: a meta-review.* World psychiatry : official journal of the World Psychiatric Association (WPA), 2014. **13**(2): p. 153-160.
9. *Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017.* Lancet, 2018. **392**(10159): p. 1736-1788.
10. Colton, C.W. and R.W. Manderscheid, *Congruencies in increased mortality rates, years of potential life lost, and causes of death among public mental health clients in eight states.* Preventing chronic disease, 2006. **3**(2): p. A42-A42.
11. Whiteford, H.A., et al., *Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010.* Lancet, 2013. **382**(9904): p. 1575-86.
12. Malzberg, B., *Mortality among patients with involution melancholia.* American Journal of Psychiatry, 1937. **93**(5): p. 1231-1238.
13. Dregan, A., et al., *Potential gains in life expectancy from reducing amenable mortality among people diagnosed with serious mental illness in the United Kingdom.* PLoS One, 2020. **15**(3): p. e0230674.
14. Laursen, T.M., M. Nordentoft, and P.B. Mortensen, *Excess early mortality in schizophrenia.* Annual review of clinical psychology, 2014. **10**: p. 425-448.
15. Druss, B.G. and E.R. Walker, *Mental disorders and medical comorbidity.* The Synthesis project. Research synthesis report, 2011(21): p. 1-26.
16. Walker, E.R., R.E. McGee, and B.G. Druss, *Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis.* JAMA psychiatry, 2015. **72**(4): p. 334-341.
17. Saha, S., D. Chant, and J. McGrath, *A systematic review of mortality in schizophrenia: is the differential mortality gap worsening over time?* Archives of general psychiatry, 2007. **64**(10): p. 1123-1131.
18. Brown, S., *Excess mortality of schizophrenia. A meta-analysis.* Br J Psychiatry, 1997. **171**: p. 502-8.
19. Osby, U., et al., *Time trends in schizophrenia mortality in Stockholm county, Sweden: cohort study.* BMJ (Clinical research ed.), 2000. **321**(7259): p. 483-484.
20. Murray, C.J. and A. Lopez, *The Global burden of disease : a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. 1996, Cambridge: Harvard University Press.
21. *Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017.* Lancet, 2018. **392**(10159): p. 1859-1922.
22. Murray, C.J. and A.D. Lopez, *Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study.* Lancet, 1997. **349**(9064): p. 1498-504.
23. Organization, W.H., *Mental health and development: targeting people with mental health conditions as a vulnerable group*, in *Mental health and development: targeting people with mental health conditions as a vulnerable group*. 2010. p. 74-74.
24. Bloom, D., et al., *The global economic burden of noncommunicable diseases.* 2011, Geneva: World Economic Forum.
25. World Health, O., *Investing in mental health*. 2003, Geneva, Switzerland: Department of Mental Health and Substance Dependence, Noncommunicable Diseases and Mental Health, World Health Organization.
26. Grover, S., et al., *Cost of care of schizophrenia: a study of Indian out-patient attenders.* Acta Psychiatr Scand, 2005. **112**(1): p. 54-63.

27. Somaiya, M., et al., *Comparative study of cost of care of outpatients with bipolar disorder and schizophrenia.* Asian J Psychiatr, 2014. **12**: p. 125-33.
28. Alonso, J., et al., *Days out of role due to common physical and mental conditions: results from the WHO World Mental Health surveys.* Mol Psychiatry, 2011. **16**(12): p. 1234-46.
29. Galatzer-Levy, I.R. and R.M. Galatzer-Levy, *The revolution in psychiatric diagnosis: problems at the foundations.* Perspect Biol Med, 2007. **50**(2): p. 161-80.
30. van Os, J. and S. Kapur, *Schizophrenia.* Lancet, 2009. **374**(9690): p. 635-45.
31. Organization, W.H., *ICD-10: The ICD-10 Classification of Mental and Behavioural Disorders: diagnostic criteria for research*, in *ICD-10: the ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. 1993.
32. Association, A.P., *Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed*. Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed. 2013, Arlington, VA, US: American Psychiatric Publishing, Inc. xliv, 947-xliv, 947.
33. Tasman, A., J. Kay, and R. Ursano, *The Psychiatric Interview: Evaluation and Diagnosis*. 2013.
34. Cantor, N. and N. Genero, *Psychiatric diagnosis and natural categorization: A close analogy*, in *Contemporary directions in psychopathology: Toward the DSM-IV.* 1986, The Guilford Press: New York, NY, US. p. 233-256.
35. Kim, N.S. and W.K. Ahn, *Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory.* J Exp Psychol Gen, 2002. **131**(4): p. 451-76.
36. Schneider, K., *Clinical psychopathology.(trans. by MW Hamilton).* 1959.
37. Goodwin, F.K. and K.R. Jamison, *Manic-depressive illness*. Manic-depressive illness. 1990, New York, NY, US: Oxford University Press. xxi, 938-xxi, 938.
38. Dunayevich, E. and P.E. Keck, Jr., *Prevalence and description of psychotic features in bipolar mania.* Curr Psychiatry Rep, 2000. **2**(4): p. 286-90.
39. Gonzalez-Pinto, A., et al., *First episode in bipolar disorder: misdiagnosis and psychotic symptoms.* J Affect Disord, 1998. **50**(1): p. 41-4.
40. Canuso, C.M., et al., *Psychotic symptoms in patients with bipolar mania.* J Affect Disord, 2008. **111**(2-3): p. 164-9.
41. Jääskeläinen, E., et al., *Epidemiology of psychotic depression - systematic review and meta-analysis.* Psychol Med, 2018. **48**(6): p. 905-918.
42. Toh, W.L., N. Thomas, and S.L. Rossell, *Auditory verbal hallucinations in bipolar disorder (BD) and major depressive disorder (MDD): A systematic review.* J Affect Disord, 2015. **184**: p. 18-28.
43. Miller, P.R., et al., *Inpatient diagnostic assessments: 1. Accuracy of structured vs. unstructured interviews.* Psychiatry Res, 2001. **105**(3): p. 255-64.
44. Phillips, M.L. and D.J. Kupfer, *Bipolar disorder diagnosis: challenges and future directions.* Lancet, 2013. **381**(9878): p. 1663-71.
45. Kaltenboeck, A., D. Winkler, and S. Kasper, *Bipolar and related disorders in DSM-5 and ICD-10.* CNS Spectr, 2016. **21**(4): p. 318-23.
46. Biedermann, F. and W.W. Fleischhacker, *Psychotic disorders in DSM-5 and ICD-11.* CNS Spectr, 2016. **21**(4): p. 349-54.
47. Santelmann, H., et al., *Interrater reliability of schizoaffective disorder compared with schizophrenia, bipolar disorder, and unipolar depression - A systematic review and meta-analysis.* Schizophr Res, 2016. **176**(2-3): p. 357-363.
48. Santelmann, H., et al., *Test-retest reliability of schizoaffective disorder compared with schizophrenia, bipolar disorder, and unipolar depression--a systematic review and meta-analysis.* Bipolar Disord, 2015. **17**(7): p. 753-68.
49. Bannister, D., *The logical requirements o f research into schizophrenia.* Br J Psychiatry, 1968. **114**(507): p. 181-8.
50. Olbert, C.M., G.J. Gala, and L.A. Tupler, *Quantifying heterogeneity attributable to polythetic diagnostic criteria: theoretical framework and empirical application.* J Abnorm Psychol, 2014. **123**(2): p. 452-62.
51. Meyer, F. and T.D. Meyer, *The misdiagnosis of bipolar disorder as a psychotic disorder: some of its causes and their influence on therapy.* J Affect Disord, 2009. **112**(1-3): p. 174-83.
52. Kessler, R.C., et al., *The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R).* JAMA, 2003. **289**(23): p. 3095-3105.
53. Newson, J.J., D. Hunter, and T.C. Thiagarajan, *The Heterogeneity of Mental Health Assessment.* Frontiers in psychiatry, 2020. **11**: p. 76-76.
54. Akiskal, H.S., et al., *Re-evaluating the prevalence of and diagnostic composition within the broad clinical spectrum of bipolar disorders.* Journal of affective disorders, 2000. **59 Suppl 1**: p. S5-S30.
55. Hirschfeld, R.M., *Bipolar spectrum disorder: improving its recognition and diagnosis.* The Journal of clinical psychiatry, 2001. **62 Suppl 14**: p. 5-9.
56. Watanabe, K., et al., *Perceptions and impact of bipolar disorder in Japan: results of an Internet survey.* Neuropsychiatr Dis Treat, 2016. **12**: p. 2981-2987.
57. Judd, L.L., et al., *Long-term symptomatic status of bipolar I vs. bipolar II disorders.* Int J Neuropsychopharmacol, 2003. **6**(2): p. 127-37.
58. Hirschfeld, R.M., L. Lewis, and L.A. Vornik, *Perceptions and impact of bipolar disorder: how far have we really come? Results of the national depressive and manic-depressive association 2000 survey of individuals with bipolar disorder.* J Clin Psychiatry, 2003. **64**(2): p. 161-74.
59. Lish, J.D., et al., *The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members.* J Affect Disord, 1994. **31**(4): p. 281-94.
60. Bruchmuller, K. and T.D. Meyer, *Diagnostically irrelevant information can affect the likelihood of a diagnosis of bipolar disorder.* J Affect Disord, 2009. **116**(1-2): p. 148-51.
61. Baldessarini, R.J., et al., *Effects of treatment latency on response to maintenance treatment in manic-depressive disorders.* Bipolar Disord, 2007. **9**(4): p. 386-93.
62. Judd, L.L., et al., *The long-term natural history of the weekly symptomatic status of bipolar I disorder.* Arch Gen Psychiatry, 2002. **59**(6): p. 530-7.
63. Hirschfeld, R.M., *Differential diagnosis of bipolar disorder and major depressive disorder.* J Affect Disord, 2014. **169 Suppl 1**: p. S12-6.
64. Fountoulakis, K.N., et al., *Mood Symptoms in Stabilized Patients with Schizophrenia: A Bipolar Type with Predominant Psychotic Features?* Psychiatr Danub, 2017. **29**(2): p. 148-154.
65. Altamura, A.C., et al., *Misdiagnosis, duration of untreated illness (DUI) and outcome in bipolar patients with psychotic symptoms: A naturalistic study.* Journal of affective disorders, 2015. **182**: p. 70-75.

66. Kennedy, S.H., et al., *Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 Clinical Guidelines for the Management of Adults with Major Depressive Disorder: Section 3. Pharmacological Treatments.* Can J Psychiatry, 2016. **61**(9): p. 540-60.
67. Yatham, L.N., et al., *Canadian Network for Mood and Anxiety Treatments (CANMAT) guidelines for the management of patients with bipolar disorder: consensus and controversies.* Bipolar Disord, 2005. **7 Suppl 3**: p. 5-69.
68. Remington, G., et al., *Guidelines for the Pharmacotherapy of Schizophrenia in Adults.* Can J Psychiatry, 2017. **62**(9): p. 604-616.
69. Awad, A.G., et al., *Quality of life among bipolar disorder patients misdiagnosed with major depressive disorder.* Prim Care Companion J Clin Psychiatry, 2007. **9**(3): p. 195-202.
70. Altshuler, L.L., et al., *Antidepressant-induced mania and cycle acceleration: a controversy revisited.* Am J Psychiatry, 1995. **152**(8): p. 1130-8.
71. Boerlin, H.L., et al., *Bipolar depression and antidepressant-induced mania: a naturalistic study.* J Clin Psychiatry, 1998. **59**(7): p. 374-9.
72. Peet, M., *Induction of mania with selective serotonin re-uptake inhibitors and tricyclic antidepressants.* Br J Psychiatry, 1994. **164**(4): p. 549-50.
73. Shi, L., P. Thiebaud, and J.S. McCombs, *The impact of unrecognized bipolar disorders for patients treated for depression with antidepressants in the fee-for-services California Medicaid (Medi-Cal) program.* J Affect Disord, 2004. **82**(3): p. 373-83.
74. Angst, F., et al., *Mortality of patients with mood disorders: follow-up over 34-38 years.* J Affect Disord, 2002. **68**(2-3): p. 167-81.
75. Tondo, L., B. Lepri, and R.J. Baldessarini, *Suicidal risks among 2826 Sardinian major affective disorder patients.* Acta Psychiatr Scand, 2007. **116**(6): p. 419-28.
76. McCombs, J.S., et al., *The impact of unrecognized bipolar disorders among patients treated for depression with antidepressants in the fee-for-services California Medicaid (Medi-Cal) program: a 6-year retrospective analysis.* Journal of affective disorders, 2007. **97**(1-3): p. 171-179.
77. Keck, P.E., Jr., R.C. Kessler, and R. Ross, *Clinical and economic effects of unrecognized or inadequately treated bipolar disorder.* Journal of psychiatric practice, 2008. **14 Suppl 2**: p. 31-38.
78. Lally, J. and J.H. MacCabe, *Antipsychotic medication in schizophrenia: a review.* British medical bulletin, 2015. **114**(1): p. 169-179.
79. Al-Harbi, K.S., *Treatment-resistant depression: therapeutic trends, challenges, and future directions.* Patient preference and adherence, 2012. **6**: p. 369-388.
80. Sharma, V., M. Khan, and A. Smith, *A closer look at treatment resistant depression: is it due to a bipolar diathesis?* Journal of affective disorders, 2005. **84**(2-3): p. 251-257.
81. Correa, R., et al., *Is unrecognized bipolar disorder a frequent contributor to apparent treatment resistant depression?* Journal of affective disorders, 2010. **127**(1-3): p. 10-18.
82. Mitelman, S.A., *Transdiagnostic neuroimaging in psychiatry: A review.* Psychiatry research, 2019. **277**: p. 23-38.
83. Aydin, O., P. Unal Aydin, and A. Arslan, *Development of Neuroimaging-Based Biomarkers in Psychiatry.* Adv Exp Med Biol, 2019. **1192**: p. 159-195.
84. Goodkind, M., et al., *Identification of a Common Neurobiological Substrate for Mental Illness.* JAMA Psychiatry, 2015. **72**(4): p. 305-315.
85. White, T., et al., *Comparative neuropsychiatry: White matter abnormalities in children and adolescents with schizophrenia, bipolar affective disorder, and obsessive-compulsive disorder.* European Psychiatry, 2015. **30**(2): p. 205-213.
86. Chang, M., et al., *Neurobiological Commonalities and Distinctions Among Three Major Psychiatric Diagnostic Categories: A Structural MRI Study.* Schizophrenia Bulletin, 2017. **44**(1): p. 65-74.
87. Smoller, J.W. and C.T. Finn, *Family, twin, and adoption studies of bipolar disorder.* Am J Med Genet C Semin Med Genet, 2003. **123c**(1): p. 48-58.
88. Barnett, J.H. and J.W. Smoller, *The genetics of bipolar disorder.* Neuroscience, 2009. **164**(1): p. 331-43.
89. Gejman, P.V., A.R. Sanders, and J. Duan, *The role of genetics in the etiology of schizophrenia.* Psychiatr Clin North Am, 2010. **33**(1): p. 35-66.
90. Cardno, A.G. and Gottesman, II, *Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics.* Am J Med Genet, 2000. **97**(1): p. 12-7.
91. Cardno, A.G., et al., *Heritability estimates for psychotic disorders: the Maudsley twin psychosis series.* Arch Gen Psychiatry, 1999. **56**(2): p. 162-8.
92. Prata, D.P., et al., *Unravelling the genetic basis of schizophrenia and bipolar disorder with GWAS: A systematic review.* J Psychiatr Res, 2019. **114**: p. 178-207.
93. Davalieva, K., I. Maleva Kostovska, and A.J. Dwork, *Proteomics Research in Schizophrenia.* Front Cell Neurosci, 2016. **10**: p. 18.
94. Graves, P.R. and T.A. Haystead, *Molecular biologist's guide to proteomics.* Microbiol Mol Biol Rev, 2002. **66**(1): p. 39-63; table of contents.
95. Bot, M., et al., *Serum proteomic profiling of major depressive disorder.* Transl Psychiatry, 2015. **5**: p. e599.
96. van Gastel, J., et al., *Enhanced Molecular Appreciation of Psychiatric Disorders Through High-Dimensionality Data Acquisition and Analytics.* Methods Mol Biol, 2019. **2011**: p. 671-723.
97. Saia-Cereda, V.M., et al., *Psychiatric disorders biochemical pathways unraveled by human brain proteomics.* European archives of psychiatry and clinical neuroscience, 2017. **267**(1): p. 3-17.
98. Comes, A.L., et al., *Proteomics for blood biomarker exploration of severe mental illness: pitfalls of the past and potential for the future.* Translational Psychiatry, 2018. **8**(1): p. 160.
99. Sokolowska, I., et al., *The potential of biomarkers in psychiatry: focus on proteomics.* J Neural Transm (Vienna), 2015. **122 Suppl 1**: p. S9-18.
100. Rollins, B., et al., *Analysis of whole genome biomarker expression in blood and brain.* Am J Med Genet B Neuropsychiatr Genet, 2010. **153b**(4): p. 919-36.
101. Guest, P.C., *Proteomic Studies of Psychiatric Disorders.* Methods Mol Biol, 2018. **1735**: p. 59-89.
102. Giusti, L., et al., *Proteomics insight into psychiatric disorders: an update on biological fluid biomarkers.* Expert Rev Proteomics, 2016. **13**(10): p. 941-950.
103. Wang, F., et al., *Characteristics of Peripheral Lymphocyte Subset Alteration in COVID-19 Pneumonia.* J Infect Dis, 2020. **221**(11): p. 1762-1769.
104. Tappuni, A.R., *Immune reconstitution inflammatory syndrome.* Adv Dent Res, 2011. **23**(1): p. 90-6.

105. Nowicka, D., et al., *NK and NKT-Like Cells in Patients with Recurrent Furunculosis.* Arch Immunol Ther Exp (Warsz), 2018. **66**(4): p. 315-319.
106. Chakrabarty, T. and L.N. Yatham, *Objective and biological markers in bipolar spectrum presentations.* Expert Rev Neurother, 2019. **19**(3): p. 195-209.
107. Schwarz, E., et al., *Identification of a biological signature for schizophrenia in serum.* Mol Psychiatry, 2012. **17**(5): p. 494-502.
108. Chan, M.K., et al., *Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset.* Translational Psychiatry, 2015. **5**(7): p. e601-e601.
109. Papakostas, G.I., et al., *Assessment of a multi-assay, serum-based biological diagnostic test for major depressive disorder: a pilot and replication study.* Mol Psychiatry, 2013. **18**(3): p. 332-9.
110. Bilello, J.A., et al., *MDDScore: confirmation of a blood test to aid in the diagnosis of major depressive disorder.* The Journal of clinical psychiatry, 2015. **76**(2): p. e199-e206.
111. Stelzhammer, V., et al., *Proteomic changes in serum of first onset, antidepressant drug-naïve major depression patients.* Int J Neuropsychopharmacol, 2014. **17**(10): p. 1599-608.
112. Haenisch, F., et al., *Towards a blood-based diagnostic panel for bipolar disorder.* Brain Behav Immun, 2016. **52**: p. 49-57.
113. Frye, M.A., et al., *Feasibility of investigating differential proteomic expression in depression: implications for biomarker development in mood disorders.* Translational psychiatry, 2015. **5**(12): p. e689-e689.
114. Domenici, E., et al., *Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections.* PLoS One, 2010. **5**(2): p. e9166.
115. Chen, J., et al., *Comparative proteomic analysis of plasma from bipolar depression and depressive disorder: identification of proteins associated with immune regulatory.* Protein & cell, 2015. **6**(12): p. 908-911.
116. Ren, J., et al., *Identification of plasma biomarkers for distinguishing bipolar depression from major depressive disorder by iTRAQ-coupled LC-MS/MS and bioinformatics analysis.* Psychoneuroendocrinology, 2017. **86**: p. 17-24.
117. Rhee, S.J., et al., *Comparison of serum protein profiles between major depressive disorder and bipolar disorder.* BMC Psychiatry, 2020. **20**(1): p. 145.
118. Teixeira, A.L., et al., *Update on bipolar disorder biomarker candidates.* Expert review of molecular diagnostics, 2016. **16**(11): p. 1209-1220.
119. Le-Niculescu, H., et al., *Identifying blood biomarkers for mood disorders using convergent functional genomics.* Molecular Psychiatry, 2009. **14**(2): p. 156-174.
120. Song, Y.R., et al., *Specific alterations in plasma proteins during depressed, manic, and euthymic states of bipolar disorder.* Braz J Med Biol Res, 2015. **48**(11): p. 973-82.
121. Končarević, S., et al., *In-depth profiling of the peripheral blood mononuclear cells proteome for clinical blood proteomics.* International journal of proteomics, 2014. **2014**: p. 129259-129259.
122. Haudek-Prinz, V.J., et al., *Proteome signatures of inflammatory activated primary human peripheral blood mononuclear cells.* Journal of proteomics, 2012. **76 Spec No.**(5): p. 150-162.
123. Corkum, C.P., et al., *Immune cell subsets and their gene expression profiles from human PBMC isolated by Vacutainer Cell Preparation Tube (CPT™) and standard density gradient.* BMC immunology, 2015. **16**: p. 48-48.
124. Grievink, H.W., et al., *Comparison of Three Isolation Techniques for Human Peripheral Blood Mononuclear Cells: Cell Recovery and Viability, Population Composition, and Cell Functionality.* Biopreserv Biobank, 2016. **14**(5): p. 410-415.
125. Ruitenberg, J.J., et al., *VACUTAINER CPT and Ficoll density gradient separation perform equivalently in maintaining the quality and function of PBMC from HIV seropositive blood samples.* BMC immunology, 2006. **7**: p. 11-11.
126. Nilsson, C., et al., *Optimal blood mononuclear cell isolation procedures for gamma interferon enzyme-linked immunospot testing of healthy Swedish and Tanzanian subjects.* Clinical and vaccine immunology : CVI, 2008. **15**(4): p. 585-589.
127. Gladkevich, A., H.F. Kauffman, and J. Korf, *Lymphocytes as a neural probe: potential for studying psychiatric disorders.* Prog Neuropsychopharmacol Biol Psychiatry, 2004. **28**(3): p. 559-76.
128. Fan, H.M., et al., *Altered microRNA Expression in Peripheral Blood Mononuclear Cells from Young Patients with Schizophrenia.* J Mol Neurosci, 2015. **56**(3): p. 562-71.
129. Fisar, Z. and J. Raboch, *Depression, antidepressants, and peripheral blood components.* Neuro Endocrinol Lett, 2008. **29**(1): p. 17-28.
130. Bhandage, A.K., et al., *GABA Regulates Release of Inflammatory Cytokines From Peripheral Blood Mononuclear Cells and CD4(+) T Cells and Is Immunosuppressive in Type 1 Diabetes.* EBioMedicine, 2018. **30**: p. 283-294.
131. Marques-Deak, A., G. Cizza, and E. Sternberg, *Brain-immune interactions and disease susceptibility.* Molecular Psychiatry, 2005. **10**(3): p. 239-250.
132. DeRijk, R., et al., *Exercise and Circadian Rhythm-Induced Variations in Plasma Cortisol Differentially Regulate Interleukin-1ß (IL-1ß), IL-6, and Tumor Necrosis Factor-α (TNFα) Production in Humans: High Sensitivity of TNFα and Resistance of IL-6.* The Journal of Clinical Endocrinology & Metabolism, 1997. **82**(7): p. 2182-2191.
133. Hornung, V., et al., *Quantitative expression of toll-like receptor 1-10 mRNA in cellular subsets of human peripheral blood mononuclear cells and sensitivity to CpG oligodeoxynucleotides.* J Immunol, 2002. **168**(9): p. 4531-7.
134. Ringheim, G.E., K.L. Burgher, and J.A. Heroux, *Interleukin-6 mRNA expression by cortical neurons in culture: evidence for neuronal sources of interleukin-6 production in the brain.* J Neuroimmunol, 1995. **63**(2): p. 113-23.
135. Sébire, G., et al., *In vitro production of IL-6, IL-1 beta, and tumor necrosis factor-alpha by human embryonic microglial and neural cells.* J Immunol, 1993. **150**(4): p. 1517-23.
136. Guyon, A., et al., *How cytokines can influence the brain: a role for chemokines?* J Neuroimmunol, 2008. **198**(1-2): p. 46-55.
137. Turnbull, A.V. and C.L. Rivier, *Regulation of the hypothalamic-pituitary-adrenal axis by cytokines: actions and mechanisms of action.* Physiol Rev, 1999. **79**(1): p. 1-71.
138. Krueger, J.M., et al., *Sleep. A physiologic role for IL-1 beta and TNF-alpha.* Ann N Y Acad Sci, 1998. **856**: p. 148-59.
139. Giulian, D., et al., *Interleukin-1 is an astroglial growth factor in the developing brain.* J Neurosci, 1988. **8**(2): p. 709-14.
140. Boulanger, L.M., *Immune proteins in brain development and synaptic plasticity.* Neuron, 2009. **64**(1): p. 93-109.
141. Quan, N. and W.A. Banks, *Brain-immune communication pathways.* Brain Behav Immun, 2007. **21**(6): p. 727-35.
142. Kokai, M., et al., *Immunophenotypic studies on atypical lymphocytes in psychiatric patients.* Psychiatry Res, 1998. **77**(2): p. 105-12.

143. Cosentino, M., et al., *Assessment of lymphocyte subsets and neutrophil leukocyte function in chronic psychiatric patients on long-term drug therapy.* Prog Neuropsychopharmacol Biol Psychiatry, 1996. **20**(7): p. 1117-29.
144. Xu, Y.Y., et al., *MicroRNA-Based Biomarkers in the Diagnosis and Monitoring of Therapeutic Response in Patients with Depression.* Neuropsychiatr Dis Treat, 2019. **15**: p. 3583-3597.
145. Zheng, P., et al., *Metabolite signature for diagnosing major depressive disorder in peripheral blood mononuclear cells.* J Affect Disord, 2016. **195**: p. 75-81.
146. Herberth, M., et al., *Peripheral profiling analysis for bipolar disorder reveals markers associated with reduced cell survival.* Proteomics, 2011. **11**(1): p. 94-105.
147. Coppens, V., et al., *Profiling of the Peripheral Blood Mononuclear Cell Proteome in Schizophrenia and Mood Disorders for the Discovery of Discriminatory Biomarkers: A Proof-of-Concept Study.* Neuropsychobiology, 2020: p. 1-11.
148. Torres, K.C., et al., *The leukocytes expressing DARPP-32 are reduced in patients with schizophrenia and bipolar disorder.* Prog Neuropsychopharmacol Biol Psychiatry, 2009. **33**(2): p. 214-9.
149. Torres, K.C., et al., *Expression of neuronal calcium sensor-1 (NCS-1) is decreased in leukocytes of schizophrenia and bipolar disorder patients.* Prog Neuropsychopharmacol Biol Psychiatry, 2009. **33**(2): p. 229-34.
150. Ferretjans, R., et al., *Cognitive performance and peripheral endocannabinoid system receptor expression in schizophrenia.* Schizophr Res, 2014. **156**(2-3): p. 254-60.
151. Perl, O., et al., *The alpha7 nicotinic acetylcholine receptor in schizophrenia: decreased mRNA levels in peripheral blood lymphocytes.* Faseb j, 2003. **17**(13): p. 1948-50.
152. Gardiner, E.J., et al., *Gene expression analysis reveals schizophrenia-associated dysregulation of immune pathways in peripheral blood mononuclear cells.* J Psychiatr Res, 2013. **47**(4): p. 425-37.
153. Gurvich, A., et al., *A role for prostaglandins in rapid cycling suggested by episode-specific gene expression shifts in peripheral blood mononuclear cells: a preliminary report.* Bipolar Disord, 2014. **16**(8): p. 881-8.
154. Roy, B., R.C. Shelton, and Y. Dwivedi, *DNA methylation and expression of stress related genes in PBMC of MDD patients with and without serious suicidal ideation.* J Psychiatr Res, 2017. **89**: p. 115-124.
155. Ignjatovic, V., et al., *Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data.* J Proteome Res, 2019.
156. Luque-Garcia, J.L. and T.A. Neubert, *Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry.* J Chromatogr A, 2007. **1153**(1-2): p. 259-76.
157. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects.* Mol Cell Proteomics, 2002. **1**(11): p. 845-67.
158. Omenn, G.S., *Exploring the human plasma proteome.* Proteomics, 2005. **5**(13): p. 3223, 3225.
159. Ramstrom, M., et al., *Depletion of high-abundant proteins in body fluids prior to liquid chromatography fourier transform ion cyclotron resonance mass spectrometry.* J Proteome Res, 2005. **4**(2): p. 410-6.
160. Echan, L.A., et al., *Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma.* Proteomics, 2005. **5**(13): p. 3292-303.
161. Geyer, P.E., et al., *Revisiting biomarker discovery by plasma proteomics.* Molecular systems biology, 2017. **13**(9): p. 942-942.
162. Huang, L., et al., *Immunoaffinity separation of plasma proteins by IgY microbeads: meeting the needs of proteomic sample preparation and analysis.* Proteomics, 2005. **5**(13): p. 3314-28.
163. Giansanti, P., et al., *Six alternative proteases for mass spectrometry-based proteomics beyond trypsin.* Nat Protoc, 2016. **11**(5): p. 993-1006.
164. Hsieh, E.J., et al., *Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples.* J Am Soc Mass Spectrom, 2013. **24**(1): p. 148-53.
165. Zhang, Y., et al., *Evaluating Chromatographic Approaches for the Quantitative Analysis of a Human Proteome on Orbitrap-Based Mass Spectrometry Systems.* J Proteome Res, 2019. **18**(4): p. 1857-1869.
166. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.
167. Weisser, H., et al., *An automated pipeline for high-throughput label-free quantitative proteomics.* J Proteome Res, 2013. **12**(4): p. 1628-44.
168. Cox, J., et al., *Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ.* Mol Cell Proteomics, 2014. **13**(9): p. 2513-26.
169. Ahad, N.A., et al., *Sensitivity of normality tests to non-normal data.* Sains Malaysiana, 2011. **40**(6): p. 637-641.
170. Nahm, F.S., *Nonparametric statistical tests for the continuous data: the basic concept and the practical use.* Korean journal of anesthesiology, 2016. **69**(1): p. 8-14.
171. Food and D. Administration, *Guidance for industry: bioanalytical method validation.* http://www.fda.gov/cder/Guidance/4252fnl.pdf,2001.
172. Maudsley, S., et al., *Bioinformatic approaches to metabolic pathways analysis.* Methods in molecular biology (Clifton, N.J.), 2011. **756**: p. 99-130.
173. Cai, H., et al., *VennPlex--a novel Venn diagram program for comparing and visualizing datasets with differentially regulated datapoints.* PLoS One, 2013. **8**(1): p. e53388.
174. Wu, D. and G.K. Smyth, *Camera: a competitive gene set test accounting for inter-gene correlation.* Nucleic Acids Res, 2012. **40**(17): p. e133.
175. Gatti, D.M., et al., *Heading down the wrong pathway: on the influence of correlation within gene sets.* BMC Genomics, 2010. **11**: p. 574.
176. Breslin, T., P. Edén, and M. Krogh, *Comparing functional annotation analyses with Catmap.* BMC Bioinformatics, 2004. **5**: p. 193.
177. Dørum, G., et al., *Rotation testing in gene set enrichment analysis for small direct comparison experiments.* Stat Appl Genet Mol Biol, 2009. **8**: p. Article34.
178. Nguyen, T.-M., et al., *Identifying significantly impacted pathways: a comprehensive review and assessment.* Genome Biology, 2019. **20**(1): p. 203.
179. Geistlinger, L., et al., *Toward a gold standard for benchmarking gene set enrichment analysis.* Briefings in Bioinformatics, 2020.
180. Maleki, F., et al., *Size matters: how sample size affects the reproducibility and specificity of gene set analysis.* Hum Genomics, 2019. **13**(Suppl 1): p. 42.

181. Maleki, F., et al., *Measuring consistency among gene set analysis methods: A systematic study.* J Bioinform Comput Biol, 2019. **17**(5): p. 1940010.
182. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Research, 2015. **43**(7): p. e47-e47.
183. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0.* Bioinformatics, 2011. **27**(12): p. 1739-1740.
184. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
185. Merico, D., et al., *Enrichment map: a network-based method for gene-set enrichment visualization and interpretation.* PLoS One, 2010. **5**(11): p. e13984.
186. Kucera, M., et al., *AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations.* F1000Res, 2016. **5**: p. 1717.
187. Oesper, L., et al., *WordCloud: a Cytoscape plugin to create a visual semantic summary of networks.* Source Code Biol Med, 2011. **6**: p. 7.
188. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.* Nucleic Acids Res, 2019. **47**(D1): p. D607-d613.
189. Doncheva, N.T., et al., *Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data.* J Proteome Res, 2019. **18**(2): p. 623-632.
190. Assenov, Y., et al., *Computing topological parameters of biological networks.* Bioinformatics, 2008. **24**(2): p. 282-4.
191. Chin, C.H., et al., *cytoHubba: identifying hub objects and sub-networks from complex interactome.* BMC Syst Biol, 2014. **8 Suppl 4**(Suppl 4): p. S11.
192. Barabási, A.-L., N. Gulbahce, and J. Loscalzo, *Network medicine: a network-based approach to human disease.* Nature Reviews Genetics, 2011. **12**(1): p. 56-68.
193. Yu, H., et al., *The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics.* PLoS Comput Biol, 2007. **3**(4): p. e59.
194. Kohler, K. and A. Ercole, *Can network science reveal structure in a complex healthcare system? A network analysis using data from emergency surgical services.* BMJ Open, 2020. **10**(2): p. e034265.
195. Karbalaei, R., et al., *Protein-protein interaction analysis of Alzheimer`s disease and NAFLD based on systems biology methods unhide common ancestor pathways.* Gastroenterol Hepatol Bed Bench, 2018. **11**(1): p. 27-33.
196. Valizadeh, R., et al., *Evaluation of involved proteins in colon adenocarcinoma: an interactome analysis.* Gastroenterol Hepatol Bed Bench, 2017. **10**(Suppl1): p. S129-s138.
197. Rezaei-Tavirani, M., et al., *Protein-Protein Interaction Network Analysis for a Biomarker Panel Related to Human Esophageal Adenocarcinoma.* Asian Pac J Cancer Prev, 2017. **18**(12): p. 3357-3363.
198. Nepusz, T., H. Yu, and A. Paccanaro, *Detecting overlapping protein complexes in protein-protein interaction networks.* Nature methods, 2012. **9**(5): p. 471-472.
199. Zhao, L., et al., *Comparative evaluation of label-free quantification strategies.* Journal of Proteomics, 2020. **215**: p. 103669.
200. Bubis, J.A., et al., *Comparative evaluation of label-free quantification methods for shotgun proteomics.* Rapid Commun Mass Spectrom, 2017. **31**(7): p. 606-612.
201. Ramus, C., et al., *Benchmarking quantitative label-free LC–MS data processing workflows using a complex spiked proteomic standard dataset.* Journal of Proteomics, 2016. **132**: p. 51-62.
202. Ramus, C., et al., *Spiked proteomic standard dataset for testing label-free quantitative software and statistical methods.* Data in Brief, 2016. **6**: p. 286-294.
203. Sticker, A., et al., *Robust summarization and inference in proteome-wide label-free quantification.* Mol Cell Proteomics, 2020.
204. Olsen, J.V., et al., *A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed.* Mol Cell Proteomics, 2009. **8**(12): p. 2759-69.
205. Xu, P., D.M. Duong, and J. Peng, *Systematical optimization of reverse-phase chromatography for shotgun proteomics.* J Proteome Res, 2009. **8**(8): p. 3944-50.
206. Kocher, T., R. Swart, and K. Mechtler, *Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides.* Anal Chem, 2011. **83**(7): p. 2699-704.
207. Gilar, M., et al., *Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high-performance liquid chromatography.* J Chromatogr A, 2004. **1061**(2): p. 183-92.
208. Hassis, M.E., et al., *Evaluating the effects of preanalytical variables on the stability of the human plasma proteome.* Anal Biochem, 2015. **478**: p. 14-22.
209. Shen, Q., et al., *Strong impact on plasma protein profiles by precentrifugation delay but not by repeated freeze-thaw cycles, as analyzed using multiplex proximity extension assays.* Clin Chem Lab Med, 2018. **56**(4): p. 582-594.
210. Addona, T.A., et al., *A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease.* Nat Biotechnol, 2011. **29**(7): p. 635-43.
211. Cao, Z., et al., *Systematic comparison of fractionation methods for in-depth analysis of plasma proteomes.* J Proteome Res, 2012. **11**(6): p. 3090-100.
212. Paczesny, S., et al., *Elafin is a biomarker of graft-versus-host disease of the skin.* Sci Transl Med, 2010. **2**(13): p. 13ra2.
213. Sahebekhtiari, N., et al., *Plasma Proteomics Analysis Reveals Dysregulation of Complement Proteins and Inflammation in Acquired Obesity-A Study on Rare BMI-Discordant Monozygotic Twin Pairs.* Proteomics Clin Appl, 2019. **13**(4): p. e1800173.
214. Hakimi, A., et al., *Assessment of reproducibility in depletion and enrichment workflows for plasma proteomics using label-free quantitative data-independent LC-MS.* Proteomics, 2014. **14**(1): p. 4-13.
215. Maccarrone, G., et al., *Proteome profiling of peripheral mononuclear cells from human blood.* Proteomics, 2013. **13**(5): p. 893-7.
216. Wang, L., et al., *Comparative proteome analysis of peripheral blood mononuclear cells in systemic lupus erythematosus with iTRAQ quantitative proteomics.* Rheumatol Int, 2012. **32**(3): p. 585-93.
217. Zhang, L., et al., *Network-based proteomic analysis for postmenopausal osteoporosis in Caucasian females.* Proteomics, 2016. **16**(1): p. 12-28.
218. Maes, E., et al., *Interindividual variation in the proteome of human peripheral blood mononuclear cells.* PLoS One, 2013. **8**(4): p. e61933.

219. Corzett, T.H., et al., *Statistical analysis of variation in the human plasma proteome.* Journal of biomedicine & biotechnology, 2010. **2010**: p. 258494-258494.
220. Sperner-Unterweger, B., *Immunological Aetiology of Major Psychiatric Disorders.* Drugs, 2005. **65**(11): p. 1493-1520.
221. Bauer, M.E. and A.L. Teixeira, *Inflammation in psychiatric disorders: what comes first?* Ann N Y Acad Sci, 2019. **1437**(1): p. 57-67.
222. Taskinen, M.R. and J. Borén, *Why Is Apolipoprotein CIII Emerging as a Novel Therapeutic Target to Reduce the Burden of Cardiovascular Disease?* Curr Atheroscler Rep, 2016. **18**(10): p. 59.
223. Ooi, Esther M.M., et al., *Apolipoprotein C-III: understanding an emerging cardiovascular risk factor.* Clinical Science, 2008. **114**(10): p. 611-624.
224. Sacks, F.M., et al., *VLDL, apolipoproteins B, CIII, and E, and risk of recurrent coronary events in the Cholesterol and Recurrent Events (CARE) trial.* Circulation, 2000. **102**(16): p. 1886-92.
225. Mendivil, C.O., et al., *Low-density lipoproteins containing apolipoprotein C-III and the risk of coronary heart disease.* Circulation, 2011. **124**(19): p. 2065-72.
226. Boiko, A.S., et al., *Apolipoprotein serum levels related to metabolic syndrome in patients with schizophrenia.* Heliyon, 2019. **5**(7): p. e02033.
227. Knöchel, C., et al., *Altered apolipoprotein C expression in association with cognition impairments and hippocampus volume in schizophrenia and bipolar disorder.* Eur Arch Psychiatry Clin Neurosci, 2017. **267**(3): p. 199-212.
228. Andersen, C.B.F., et al., *Haptoglobin.* Antioxid Redox Signal, 2017. **26**(14): p. 814-831.
229. Maes, M., et al., *Haptoglobin polymorphism and schizophrenia: genetic variation on chromosome 16.* Psychiatry Res, 2001. **104**(1): p. 1-9.
230. Cooper, J.D., et al., *Schizophrenia-risk and urban birth are associated with proteomic changes in neonatal dried blood spots.* Transl Psychiatry, 2017. **7**(12): p. 1290.
231. Yee, J.Y., et al., *Peripheral blood gene expression of acute phase proteins in people with first episode psychosis.* Brain Behav Immun, 2017. **65**: p. 337-341.
232. Maes, M., et al., *Acute phase proteins in schizophrenia, mania and major depression: modulation by psychotropic drugs.* Psychiatry Res, 1997. **66**(1): p. 1-11.
233. Wan, C., et al., *Abnormal changes of plasma acute phase proteins in schizophrenia and the relation between schizophrenia and haptoglobin (Hp) gene.* Amino Acids, 2007. **32**(1): p. 101-8.
234. Yang, Y., et al., *Altered levels of acute phase proteins in the plasma of patients with schizophrenia.* Anal Chem, 2006. **78**(11): p. 3571-6.
235. Maes, M., et al., *Haptoglobin phenotypes and gene frequencies in unipolar major depression.* Am J Psychiatry, 1994. **151**(1): p. 112-6.
236. Giusti, L., et al., *Search for peripheral biomarkers in patients affected by acutely psychotic bipolar disorder: a proteomic approach.* Molecular BioSystems, 2014. **10**(6): p. 1246-1254.
237. Zanetti, M., *Cathelicidins, multifunctional peptides of the innate immunity.* J Leukoc Biol, 2004. **75**(1): p. 39-48.
238. Yang, J., et al., *An iron delivery pathway mediated by a lipocalin.* Mol Cell, 2002. **10**(5): p. 1045-56.
239. Kozłowska, E., A. Wysokiński, and E. Brzezińska-Błaszczyk, *Serum levels of peptide cathelicidin LL-37 in elderly patients with depression.* Psychiatry Res, 2017. **255**: p. 156-160.
240. Kozłowska, E., et al., *Circulating cathelicidin LL-37 level is increased in euthymic patients with bipolar disorder.* J Clin Neurosci, 2018. **48**: p. 168-172.
241. Wei, L., et al., *Elevation of plasma neutrophil gelatinase-associated lipocalin (NGAL) levels in schizophrenia patients.* J Affect Disord, 2018. **226**: p. 307-312.
242. Naudé, P.J.W., et al., *Neutrophil gelatinase-associated lipocalin: A novel inflammatory marker associated with late-life depression.* Journal of Psychosomatic Research, 2013. **75**(5): p. 444-450.
243. Kozlowska, E., et al., *Human cathelicidin LL-37 - Does it influence the homeostatic imbalance in mental disorders?* J Biosci, 2018. **43**(2): p. 321-327.
244. Sternlicht, M.D. and Z. Werb, *How matrix metalloproteinases regulate cell behavior.* Annu Rev Cell Dev Biol, 2001. **17**: p. 463-516.
245. Rodríguez, D., C.J. Morrison, and C.M. Overall, *Matrix metalloproteinases: what do they not do? New substrates and biological roles identified by murine models and proteomics.* Biochim Biophys Acta, 2010. **1803**(1): p. 39-54.
246. McClellan, S.A., et al., *Matrix Metalloproteinase-9 Amplifies the Immune Response to Pseudomonas aeruginosa Corneal Infection.* Investigative Ophthalmology & Visual Science, 2006. **47**(1): p. 256-264.
247. Lepeta, K. and L. Kaczmarek, *Matrix Metalloproteinase-9 as a Novel Player in Synaptic Plasticity and Schizophrenia.* Schizophrenia bulletin, 2015. **41**(5): p. 1003-1009.
248. Rybakowski, J.K., et al., *Matrix metalloproteinase-9 gene and bipolar mood disorder.* Neuromolecular Med, 2009. **11**(2): p. 128-32.
249. Gao, J., et al., *DNA Methylation and Gene Expression of Matrix Metalloproteinase 9 Gene in Deficit and Non-deficit Schizophrenia.* Front Genet, 2018. **9**: p. 646.
250. Kumarasinghe, N., et al., *Gene expression profiling in treatment-naive schizophrenia patients identifies abnormalities in biological pathways involving AKT1 that are corrected by antipsychotic medication.* Int J Neuropsychopharmacol, 2013. **16**(7): p. 1483-503.
251. Rybakowski, J.K., et al., *Functional polymorphism of the matrix metalloproteinase-9 (MMP-9) gene in schizophrenia.* Schizophrenia research, 2009. **109**(1-3): p. 90-93.
252. Heizmann, C.W., G. Fritz, and B.W. Schäfer, *S100 proteins: structure, functions and pathology.* Front Biosci, 2002. **7**: p. d1356-68.
253. White, S.H., W.C. Wimley, and M.E. Selsted, *Structure, function, and membrane integration of defensins.* Current opinion in structural biology, 1995. **5**(4): p. 521-527.
254. Iavarone, F., et al., *Characterization of salivary proteins of schizophrenic and bipolar disorder patients by top-down proteomics.* J Proteomics, 2014. **103**: p. 15-22.
255. Pacifico, R. and R.L. Davis, *Transcriptome sequencing implicates dorsal striatum-specific gene network, immune response and energy metabolism pathways in bipolar disorder.* Molecular Psychiatry, 2017. **22**(3): p. 441-449.
256. Craddock, R.M., et al., *Increased alpha-defensins as a blood marker for schizophrenia susceptibility.* Mol Cell Proteomics, 2008. **7**(7): p. 1204-13.
257. Lee, J., et al., *Proteomic analysis of serum from patients with major depressive disorder to compare their depressive and remission statuses.* Psychiatry investigation, 2015. **12**(2): p. 249-259.

258. Huang, T.-L., et al., *Rapid and simple analysis of disease-associated biomarkers of Taiwanese patients with schizophrenia using matrix-assisted laser desorption ionization mass spectrometry.* Clinica Chimica Acta, 2017. **473**: p. 75-81.
259. Stuart, M.J. and B.T. Baune, *Chemokines and chemokine receptors in mood disorders, schizophrenia, and cognitive impairment: a systematic review of biomarker studies.* Neurosci Biobehav Rev, 2014. **42**: p. 93-115.
260. Reale, M., et al., *Dysregulation of chemo-cytokine production in schizophrenic patients versus healthy controls.* BMC Neurosci, 2011. **12**: p. 13.
261. Grassi-Oliveira, R., et al., *Peripheral chemokine levels in women with recurrent major depression with suicidal ideation.* Revista Brasileira de Psiquiatria, 2012. **34**(1): p. 71-75.
262. van Zuiden, M., et al., *Cytokine production by leukocytes of military personnel with depressive symptoms after deployment to a combat-zone: a prospective, longitudinal study.* PLoS One, 2011. **6**(12): p. e29142.
263. Ferrúa, C.P., et al., *MicroRNAs expressed in depression and their associated pathways: A systematic review and a bioinformatics analysis.* Journal of Chemical Neuroanatomy, 2019. **100**: p. 101650.
264. Timberlake, M.A. and Y. Dwivedi, *Altered expression of endoplasmic reticulum stress associated genes in hippocampus of learned helpless rats: relevance to depression pathophysiology.* Frontiers in pharmacology, 2016. **6**: p. 319.
265. Zhang, G., et al., *Weighted Gene Coexpression Network Analysis Identifies Specific Modules and Hub Genes Related to Major Depression.* Neuropsychiatr Dis Treat, 2020. **16**: p. 703-713.
266. Perez, J., et al., *Protein kinase A and Rap1 levels in platelets of untreated patients with major depression.* Mol Psychiatry, 2001. **6**(1): p. 44-9.
267. González-Castro, T.B., et al., *Identification of gene ontology and pathways implicated in suicide behavior: Systematic review and enrichment analysis of GWAS studies.* Am J Med Genet B Neuropsychiatr Genet, 2019. **180**(5): p. 320-329.
268. Wang, H., et al., *Identification of diagnostic markers for major depressive disorder by cross-validation of data from whole blood samples.* PeerJ, 2019. **7**: p. e7171.
269. Kao, C.-F., et al., *Enriched pathways for major depressive disorder identified from a genome-wide association study.* International Journal of Neuropsychopharmacology, 2012. **15**(10): p. 1401-1411.
270. Ji, H.-F., Q.-S. Zhuang, and L. Shen, *Genetic overlap between type 2 diabetes and major depressive disorder identified by bioinformatics analysis.* Oncotarget, 2016. **7**(14): p. 17410-17414.
271. Wakabayashi, Y., et al., *State-dependent changes in the expression levels of NCAM-140 and L1 in the peripheral blood cells of bipolar disorders, but not in the major depressive disorders.* Progress in Neuro-Psychopharmacology and Biological Psychiatry, 2008. **32**(5): p. 1199-1205.
272. Gnanapavan, S. and G. Giovannoni, *Neural cell adhesion molecules in brain plasticity and disease.* Multiple sclerosis and related disorders, 2013. **2**(1): p. 13-20.
273. Sandi, C. and R. Bisaz, *A model for the involvement of neural cell adhesion molecules in stress-related mood disorders.* Neuroendocrinology, 2007. **85**(3): p. 158-176.
274. Carlson, P.J., et al., *Neural circuitry and neuroplasticity in mood disorders: insights for novel therapeutic targets.* NeuroRx, 2006. **3**(1): p. 22-41.
275. Goto, Y., C.R. Yang, and S. Otani, *Functional and dysfunctional synaptic plasticity in prefrontal cortex: roles in psychiatric disorders.* Biological psychiatry, 2010. **67**(3): p. 199-207.
276. Meyer-Lindenberg, A. and H. Tost, *Neuroimaging and plasticity in schizophrenia.* Restorative neurology and neuroscience, 2014. **32**(1): p. 119-127.
277. Wang, Q., et al., *The CMYA5 gene confers risk for both schizophrenia and major depressive disorder in the Han Chinese population.* The World Journal of Biological Psychiatry, 2014. **15**(7): p. 553-560.
278. Chen, X., et al., *GWA study data mining and independent replication identify cardiomyopathy-associated 5 (CMYA5) as a risk gene for schizophrenia.* Mol Psychiatry, 2011. **16**(11): p. 1117-29.
279. Li, M., et al., *A common variant of the cardiomyopathy associated 5 gene (CMYA5) is associated with schizophrenia in Chinese population.* Schizophr Res, 2011. **129**(2-3): p. 217-9.
280. Correll, C.U., et al., *Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls.* World psychiatry : official journal of the World Psychiatric Association (WPA), 2017. **16**(2): p. 163-180.
281. Wu, Q. and J.M. Kling, *Depression and the Risk of Myocardial Infarction and Coronary Death: A Meta-Analysis of Prospective Cohort Studies.* Medicine (Baltimore), 2016. **95**(6): p. e2815.
282. Gan, Y., et al., *Depression and the risk of coronary heart disease: a meta-analysis of prospective cohort studies.* BMC psychiatry, 2014. **14**: p. 371-371.
283. Dong, J.Y., et al., *Depression and risk of stroke: a meta-analysis of prospective studies.* Stroke, 2012. **43**(1): p. 32-7.
284. Prieto, M.L., et al., *Risk of myocardial infarction and stroke in bipolar disorder: a systematic review and exploratory meta-analysis.* Acta psychiatrica Scandinavica, 2014. **130**(5): p. 342-353.
285. Fan, Z., et al., *Schizophrenia and the risk of cardiovascular diseases: a meta-analysis of thirteen cohort studies.* J Psychiatr Res, 2013. **47**(11): p. 1549-56.
286. Li, M., et al., *Schizophrenia and risk of stroke: a meta-analysis of cohort studies.* Int J Cardiol, 2014. **173**(3): p. 588-90.
287. Yamagata, H., et al., *Identification of commonly altered genes between in major depressive disorder and a mouse model of depression.* Sci Rep, 2017. **7**(1): p. 3044.
288. *Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways.* Nat Neurosci, 2015. **18**(2): p. 199-209.
289. Saavedra, K., et al., *Epigenetic Modifications of Major Depressive Disorder.* Int J Mol Sci, 2016. **17**(8).
290. Mill, J., et al., *Epigenomic profiling reveals DNA-methylation changes associated with major psychosis.* American journal of human genetics, 2008. **82**(3): p. 696-711.
291. Connor, C.M. and S. Akbarian, *DNA methylation changes in schizophrenia and bipolar disorder.* Epigenetics, 2008. **3**(2): p. 55-8.
292. Chiu, C.-C., et al., *Polyunsaturated fatty acid deficit in patients with bipolar mania.* European Neuropsychopharmacology, 2003. **13**(2): p. 99-103.
293. Ranjekar, P.K., et al., *Decreased antioxidant enzymes and membrane essential polyunsaturated fatty acids in schizophrenic and bipolar mood disorder patients.* Psychiatry research, 2003. **121**(2): p. 109-122.

294. McNamara, R.K., et al., *Deficits in docosahexaenoic acid and associated elevations in the metabolism of arachidonic acid and saturated fatty acids in the postmortem orbitofrontal cortex of patients with bipolar disorder.* Psychiatry research, 2008. **160**(3): p. 285-299.
295. Wu, X., et al., *The comparison of glycometabolism parameters and lipid profiles between drug-naïve, first-episode schizophrenia patients and healthy controls.* Schizophrenia Research, 2013. **150**(1): p. 157-162.
296. Sengupta, S., et al., *Are metabolic indices different between drug-naïve first-episode psychosis patients and healthy controls?* Schizophrenia Research, 2008. **102**(1-3): p. 329-336.
297. Gautam, S. and P.S. Meena, *Drug-emergent metabolic syndrome in patients with schizophrenia receiving atypical (second-generation) antipsychotics.* Indian J Psychiatry, 2011. **53**(2): p. 128-33.
298. de Jesus, J.R., et al., *Simplifying the human serum proteome for discriminating patients with bipolar disorder of other psychiatry conditions.* Clin Biochem, 2017. **50**(18): p. 1118-1125.
299. Hodge, K., et al., *Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS.* J Proteomics, 2013. **88**: p. 92-103.
300. Goh, K.I., et al., *The human disease network.* Proc Natl Acad Sci U S A, 2007. **104**(21): p. 8685-90.
301. Hartwell, L.H., et al., *From molecular to modular cell biology.* Nature, 1999. **402**(6761): p. C47-C52.
302. Rolland, T., et al., *A proteome-scale map of the human interactome network.* Cell, 2014. **159**(5): p. 1212-1226.
303. Laske, C., et al., *Identification of a blood-based biomarker panel for classification of Alzheimer's disease.* International Journal of Neuropsychopharmacology, 2011. **14**(9): p. 1147-1155.
304. Pinto, J.V., et al., *Peripheral biomarker signatures of bipolar disorder and schizophrenia: A machine learning approach.* Schizophr Res, 2017. **188**: p. 182-184.
305. Jacobs, R., et al., *Identification of novel host biomarkers in plasma as candidates for the immunodiagnosis of tuberculosis disease and monitoring of tuberculosis treatment response. Oncotarget 7, 57581–57592.* 2016.
306. Lodes, M.J., et al., *Detection of cancer with serum miRNAs on an oligonucleotide microarray.* PLoS One, 2009. **4**(7): p. e6229.
307. Best, M.G., et al., *RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics.* Cancer Cell, 2015. **28**(5): p. 666-676.
308. Weatherall, M., et al., *Distinct clinical phenotypes of airways disease defined by cluster analysis.* European Respiratory Journal, 2009. **34**(4): p. 812-818.
309. Ness, R.B., et al., *A cluster analysis of bacterial vaginosis–associated microflora and pelvic inflammatory disease.* American journal of epidemiology, 2005. **162**(6): p. 585-590.
310. Erro, R., et al., *The heterogeneity of early Parkinson's disease: a cluster analysis on newly diagnosed untreated patients.* PloS one, 2013. **8**(8).
311. Newcomer, S.R., J.F. Steiner, and E.A. Bayliss, *Identifying subgroups of complex patients with cluster analysis.* The American journal of managed care, 2011. **17**(8): p. e324-32.

8. List of Abbreviations

2D-GE	2D-Gel Electrophoresis
APA	American Psychiatric Association
APOC3	APOlipoprotein C-III
AUC	Area Under the Curve
BD	Bipolar Disorder
BD-D	patients with Bipolar Disorder in a Depressive state
BD-M	patients with Bipolar Disorder in a Manic state
BMI	Body Mass Index
CAMERA	Correlation Adjusted MEan RAnk gene set test
CAMP	Cathelicidin AntiMicrobial Peptide
CCL5	C-C chemokine 5
CMYA5	CardioMYopathy Associated 5
CSF	CerebroSpinal Fluid
CV	Coefficient of Variation
DALY	Disability-Adjusted Life Year
DEFA1	neutrophil defensin 1
DEP	Differentially Expressed Protein
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ELISA	Enzyme-Linked Immuno Sorbent Assay
F	Female
GABA	Gamma-AminoButyric Acid
GBD	Global Burden of Disease
GWAS	Genome-Wide Association Studies
HC	Healthy Control
HDRS	Hamilton Depression Rating Scale
HLA-B7	HLA class I histocompatibility antigen, B-7 alpha chain
HPPP	Human Plasma Proteome Project
ICD-10	International Statistical Classification of Diseases and Related Health Problems, 10th Edition
iTRAQ	Isobaric mass-Tag labelling for Relative and Absolute Quantitation
LC	Liquid Chromatography
LCN2	neutrophil gelatinase-associated lipocalin
LCMS	Liquid Chromatography – Mass Spectrometry
M	Male
MS	Mass Spectrometry
MDD	Major Depressive Disorder
MMP9	Matrix MetalloProteinase-9
N	Sample size
NA	Not Applicable
PANSS	Positive And Negative Syndrome Scale
PANSS-P	Positive scale of the Positive And Negative Syndrome Scale
PBMC	Peripheral Blood Mononuclear Cell
PTM	Post-Translation Modifications
S100A12	protein S100A12
SD	Standard Deviation
SNP	Single-Nucleotide Polymorphism
SZ	SchiZophrenia
WHO	World Health Organization
WMH	World Mental Health
YLD	Years Lost due to Disability

YLL	Year of Life Lost
YMRS	Young Mania Rating Scale

Supplementary Materials

1. Visual Representation of Experimental Procedures

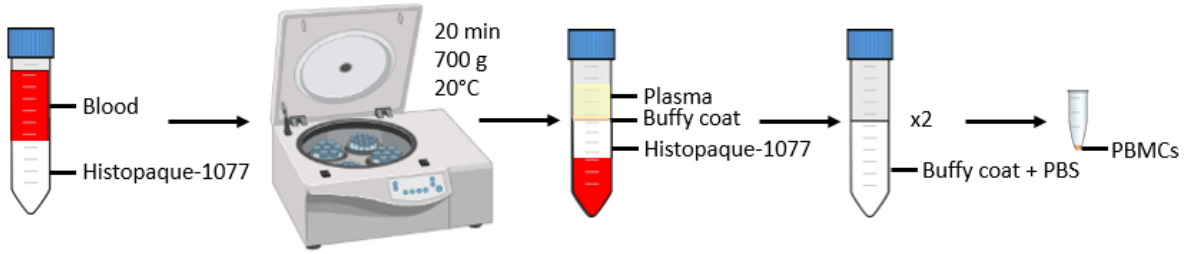

Figure S1 Visual representation of the extraction of PBMCs from whole blood.
Abbreviations: min, minutes; PBS, phosphate buffered saline; PBMCs, peripheral blood mononuclear cells

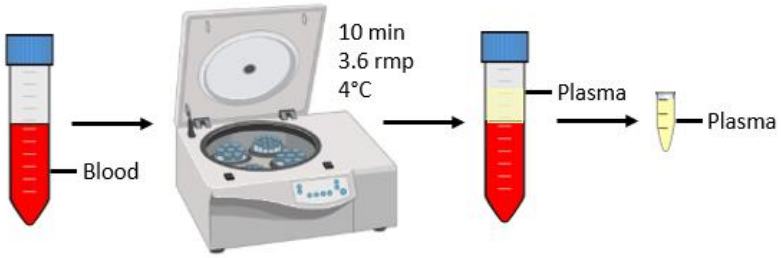

Figure S2 Visual representation of the extraction of plasma from whole blood.
Abbreviations: min, minutes; rpm, rounds per minute

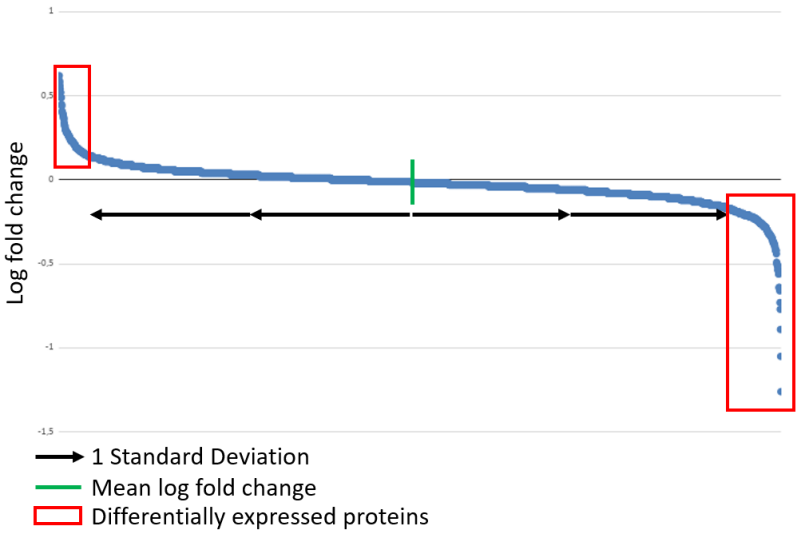

Figure S3 Visual representation of the selection of DEPs.
Abbreviations: DEPs, differentially expressed proteins

2. MaxQuant Parameters

Table S1 Settings of MaxQuant parameters

Group-specific parameters		Global parameters	
Type		**Protein quantification**	
Type	Standard	Label min. ratio count	2
Modifications		Peptides for quantification	Unique + razor
Variable modifications	Oxidation (M); Acetyl (Protein N-term); Phospho (STY)	Use only unmodified peptides and …	Yes
Fixed modifications	Carbamidomethyl (c)	Modifications used in protein quantification	Oxidation (M); Acetyl (Protein N-term); Carbamidomethyl (c)
Max. number of modifications per peptide	5	Discard unmodified counterpart peptides	Yes
Instrument		Advanced ratio estimation	Yes
Instrument type	Orbitrap	**MS/MS analyser**	
First search peptide tolerance	20	FTMS MS/MS match tolerance	20
Main search peptide tolerance	4,5	FTMS MS/MS match tolerance unit	ppm
Peptide tolerance unit	ppm	FTMS MS/MS de novo tolerance	10
Individual peptide mass tolerance	Yes	FTMS MS/MS de novo tolerance unit	ppm
Isotope match tolerance	2	FTMS MS/MS deisotoping tolerance	7
Isotope match tolerance unit	ppm	FTMS MS/MS deisotoping tolerance unit	ppm
Centroid match tolerance	8	FTMS top peaks per Da interval	12
Centroid half width	35	FTMS top x mass window (Da)	100
Centroid half width unit	ppm	FTMS de-isotoping	Yes
Time valley factor	1,4	FTMS higher charges	Yes
Isotope valley factor	1,2	FTMS water loss	Yes
Isotope time correlation	0,6	FTMS ammonia loss	Yes
Theoretical isotope correlation	0,6	FTMS dependent losses	Yes
Recalibration unit	ppm	FTMS recalibration	No
Use MS1 centroids	No	ITMS MS/MS match tolerance	0,5
Use MS2 centroids	No	ITMS MS/MS match tolerance unit	Da
Intensity dependent calibration	No	ITMS MS/MS de novo tolerance	0,25
Min. peak length	2	ITMS MS/MS de novo tolerance unit	Da
Max. charge	7	ITMS MS/MS deisotoping tolerance	0,15
Min. score for recalibration	70	ITMS MS/MS deisotoping tolerance unit	Da
Cut peaks	Yes	ITMS top peaks per Da interval	8
Gap scans	1	ITMS top x mass window (Da)	100

Advanced peak splitting	No	ITMS de-isotoping	No
Intensity threshold	0	ITMS higher charges	Yes
Check mass deficit	Yes	ITMS water loss	Yes
Intensity determination	Value at maximum	ITMS ammonia loss	Yes
Digestion		ITMS dependent losses	Yes
Digestion mode	Specific	ITMS recalibration	No
Enzyme	Trypsin/P	TOF MS/MS match tolerance	40
Max. missed	2	TOF MS/MS match tolerance unit	ppm
Label-free quantification		TOF MS/MS de novo tolerance	0,02
Label-free quantification	LFQ	TOF MS/MS de novo tolerance unit	Da
LFQ min. ratio. count	2	TOF MS/MS deisotoping tolerance	0,01
Fast LFQ	Yes	TOF MS/MS deisotoping tolerance unit	Da
LFQ min. number of neighbours	3	TOF top peaks per Da interval	10
LFQ average number of neighbours	6	TOF top x mass window (Da)	100
Skip normalization	No	TOF de-isotoping	Yes
		TOF higher charges	Yes
		TOF water loss	Yes
		TOF ammonia loss	Yes
		TOF dependent losses	Yes
		TOF recalibration	No
		Unknown MS/MS match tolerance	0,5
		Unknown MS/MS match tolerance unit	Da
		Unknown MS/MS de novo tolerance	0,25
		Unknown MS/MS de novo tolerance unit	Da
		Unknown MS/MS deisotoping tolerance	0,15
		Unknown MS/MS deisotoping tolerance unit	Da
		Unknown top peaks per Da interval	8
		Unknown top x mass window (Da)	100
		Unknown de-isotoping	No
		Unknown higher charges	Yes
		Unknown water loss	Yes
		Unknown ammonia loss	Yes
		Unknown dependent losses	Yes
		Unknown recalibration	No
		Advanced	
		Calculate peak properties	No

		Decoy mode	Revert
		Use for occupancies	Normalized ratios
		Epsilon score for mutations	Yes
		Evaluate variant peptides separately	Yes
		Disable MD5	No
		Max mods in site table	3
		Andromeda cache size	350000
		Use series reporters	No
		MS2 precursor mass shift	0
		Complementary ion ppm	20
		Identification	
		PSM FDR	0,01
		Protein FDR	0,01
		Site decoy fraction	0,01
		Min. peptides	1
		Min. razor + unique peptides	1
		Min. unique peptides	0
		Min. score for unmodified peptides	0
		Min. score for modified peptides	40
		Min. delta score for unmodified peptides	0
		Min. delta score for modified peptides	6
		Main search max. combinations	200
		Base FDR calculations on delta score	No
		Razor peptide FDR	Yes
		Second peptides	Yes
		Match between runs	Yes
		Match time window (min)	0,7
		Match ion mobility window	0,05
		Alignment time window (min)	20
		Alignment ion mobility	1
		Match unidentified features	No
		Dependent peptides	No
		Label-free quantification	
		Separate LFQ in parameter groups	No
		Stabilize large LFQ ratios	Yes
		Require MS/MS for LFQ comparisons	Yes
		iBAQ	No
		Advanced site intensities	Yes

3. Original Manuscript

PROFILING OF THE PERIPHERAL BLOOD MONONUCLEAR CELL PROTEOME IN SCHIZOPHRENIA AND MOOD DISORDERS FOR DISCOVERY OF DISCRIMINATORY BIOMARKERS - A PROOF OF CONCEPT STUDY

Coppens V.*1,2, De Wachter O.1,2, Goossens J.1,2, Hendrix J.1, Maudsley S.3, Azmi A.4, van Gastel J.3, Van Saet A.1,2, Lauwers T.1,2, Morrens M.1,2

1 Scientific Initiative for Neuropsychiatric and Psychopharmacological Studies (SINAPS), Collaborative Antwerp Psychiatric Research Institute (CAPRI), University of Antwerp, Antwerp, Belgium

2 Scientific Initiative for Neuropsychiatric and Psychopharmacological Studies (SINAPS), University Psychiatric Center Duffel, Duffel, Belgium

3 Receptor Biology Lab, Department of Biomedical Sciences, University of Antwerp, Belgium.

4 Center for Molecular Neurology, VIB, Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium.

Short Title: Peripheral blood mononuclear cell proteomics for diagnostic biomarker discovery in psychiatry

*Corresponding Author
Violette Coppens
Scientific Initiative for Neuropsychiatric and Psychopharmacological Studies (SINAPS), Collaborative Antwerp Psychiatric Research Institute (CAPRI), University of Antwerp, Antwerp, Belgium
Universiteitsplein 1, Campus Drie Eiken
Universiteitsplein 1
2610 Wilrijk

Tel: +32473556458 E-mail: violette.coppens@uantwerpen.be

Keywords: PBMC, schizophrenia, depression, biomarker, diagnosis

Abstract

Introduction: Current diagnoses in psychiatry are solely based on the evaluation of clinical presentation by the treating psychiatrist. This results in a high percentage of misdiagnosis and consequential inefficient treatment; especially with regard to major depressive disorder (MDD), depression in context of bipolar depression (BD-D), bipolar disorder with manic symptoms (BD-M) and psychosis in the context of schizophrenia (SZ). Objective biomarkers allowing for efficient discriminatory diagnostics are therefore urgently needed.

Methods: Peripheral blood mononuclear cell (PBMC) proteomes of patients with MDD (n= 5) , BD-D (n= 3), BD-M (n= 4) and SZ (n =4) and of healthy controls (HC; n = 6) were analyzed by state-of-the-art mass spectrometry. Proteins with a differential expression of >2 standard deviation (SD) expression fold change from HC and between either MDD vs. BD-D or BD-M vs. SZ were subsequently identified as potential discriminatory biomarkers.

Results: In total, 4271 individual proteins were retrieved from HC. Of these, about 2800 were detected in all patient and HC samples. For objective discrimination between MDD and BD-D, 66 candidate biomarkers were found. In parallel, 72 proteins might harbour biomarker capacity for differential diagnostics of BD-M and SZ. A single biomarker was contraregulated vs. HC in each pair of comparisons.

Discussion / Conclusion: With this work, we provide a register of candidate biomarkers with potential capacity to objectively discriminate MDD from BD-D and BD-M from SZ. Although concerning a proof-of-concept study with limited sample size, these data provide a stepping-stone for follow-up research on the validation of true discriminatory potential and feasibility of clinical implementation of the discovered biomarker candidates.

Introduction

Psychotic and mood disorders are amongst the most prevalent and debilitating psychiatric illnesses. Partial overlap in clinical presentation of these disorders often renders diagnostic differentiation between these illnesses problematic to virtually unfeasible. Illustratively, depressive episodes occur both in context of unipolar depression (major depressive disorder, MDD) and of bipolar depression (bipolar disorder, BD) with studies reporting the presence of an actual diagnosis of BD-D in 31% to up to 69% of patients misdiagnosed with unipolar depression [1,2]. Often, also phenotypical distinction between manic and psychotic symptoms remains cumbersome[3]. Clinically utilized but arguably less valid intermediate disease types such as schizoaffective disorder[4], mood disorder with psychotic symptoms, schizophrenia with depressive episodes,... and high levels of comorbidity in psychiatric disorders [5] further confound accurate diagnostics. As the different syndromes each require a specific therapeutic strategy, erroneous diagnoses may lead to severely increased and/or prolonged patient suffering. Antidepressant monotherapy for example, is in general relatively ineffective for treating bipolar depression[6]. Consequently, treatment guidelines for this population advise the primary administration of antipsychotics and mood stabilizers all or not in combination with an antidepressant [7]. Unfortunately, Viktorin et al.[8] demonstrate that 35% of bipolar patients in Sweden were treated with antidepressant monotherapy with the risk of switching to mania only occurring in these patients, while being absent in patients treated with antidepressant add-on to a mood stabilizer.

In parallel, also when suffering from comorbid psychotic symptoms bipolar patients are often misdiagnosed. A recent study reports numbers as high as 61% of BD patients with psychotic symptoms to receive an initial other diagnosis, with 21% of misdiagnoses concerning a subtype of schizophrenia [9]. Although antipsychotics (AP) are in general effective in treatment of mania [10], the resulting absence of any type of mood stabilizer in BD-M patients treated with AP can result in a substantial increase in the duration of untreated illness of bipolar patients and frequent relapses or rapid cycling [11].

The above illustrates the cruciality of accurate diagnostics and underscores the urgent need for biomarkers enabling objective discriminatory diagnosis of unipolar vs. bipolar depression and of manic vs. psychotic episodes. Although several studies have investigated the potential of different types of biomarkers to categorize patients as either MDD or BD-D, so far, no studies have successfully identified a clinically usable,

highly predictive biomarker using hypothesis-driven, targeted approaches (for review see Goossens et al., subm. and [12]). Non-targeted biomarker identification avenues (the so-called -omics approaches) however, might reveal novel molecules with high discriminatory potential that have thus far not been related to any of the investigated pathologies.

A biomarker qualifies as clinically useful if it is easy to obtain, allows for stable and accurate detection and is present in high quantities in at least 1 of the comparative groups. As peripheral blood mononuclear cells (PBMC) are effortlessly collectable and intracellular protein levels are less fluctuating than those of free-circulating plasma/serum proteins [13–16], mapping of the PBMC proteome of MDD vs. BD-D and BD-M vs. SZ patients might prove a promising strategy for novel discriminatory biomarker discovery.

Therefore, this project aims to elucidate discriminatory biomarkers for unipolar vs. bipolar depressive, manic and psychotic symptoms in context of MDD, BD and SZ by performing large-scale non-targeted liquid chromatography - mass spectrometry (LCMS) proteomics on PBMC of patients with either of the aforementioned disorders.

Methods

Patient selection

MDD patients (n=5), BD patients in a depressed state (BD-D; n=3), BD patients in a manic state (BD-M; n=4), schizophrenic patients with active positive symptoms (n=4) were recruited from the Psychiatric Hospital Duffel. In addition, 6 age- and gender matched healthy controls (HC) were recruited via advertising. Diagnosis was made according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) and confirmed by the Mini international neuropsychiatric interview (MINI) version 5.0.0. Inclusion criteria were: men and women between the ages of 18 and 55; for patients: a score of ≥17 or more on the Hamilton Depression Rating Scale (HDRS) for depressed patients (both MDD and BD-D); a score of ≥13 on the Young Mania Rating Scale (YMRS) for manic patients; a total score of ≥14 on the positive scale of the Positive and Negative Syndrome Scale (PANSS) with either a score of ≥5 on at least 1 item or a score of ≥4 on at least 2 psychotic items (P2, P3, P5). Exclusion criteria were: recent occurence or a history of chronic inflammatory disorders, autoimmune diseases, acute physical diseases and substance dependence (in the last 6 months) and

additionally for HC: a history of psychiatric disorders and having a first-degree relative with a history of psychopathology. All clinical scales were assessed by trained personnel and patients were matched to controls for age, gender, smoking status and body mass index.

The study was approved by the local ethics committee. All participants gave their written consent to take part in the study.

Collection of PBMC

Venous blood was collected sober between 7:30 and 10 AM in EDTA-coated collection tubes. Gradient centrifugation was performed using Histopaque-1077 (Sigma-aldrich, Missouri, USA) for 20 minutes at 700g and room temperature without brakes. Subsequently, the buffy coat was collected and washed 2 times with phosphate buffered saline (PBS;Thermo Fisher Scientific, Perth, United Kingdom). Finally, the supernatant was discarded en cells were stored dry at -80°C until LCMS analysis.

Quantitative Proteomics

Multiplexed iTRAQ (isobaric mass-tag labeling for relative and absolute quantitation) mass spectrometry liquid chromatography (LCMS) was performed as described hereafter. PBMC samples were solubilized in a protein extraction buffer (composition: 8 M urea, 2 M thiourea, 0.1% SDS and 50 mM triethylammonium bicarbonate). Next, protein concentrations were quantified using RC DC protein assays (Bio-Rad; California, USA). Equal amounts of proteins from each sample were then reduced by tris-2-carboxyethyl phosphine and alkylated by 5-methyl-methanoethiosulphate and finally subjected to trypsin digestion. The resulting peptides from each sample were labelled using iTRAQ reagents (Sciex, Massachusetts, USA) following the manufacturer's instructions. PBMC samples of HC and patient samples were then spread randomly across three different octaplex iTRAQ LC runs. To improve LC-MS/MS proteome coverage, samples were subjected to a 2D-LC fractionation system (Dionex ULTIMATE 3000, ThermoScientific, Massachusetts, USA). Peptide mixes were fractionated on a strong cationic exchange chromatography column (1 mm x 150 mm polysulfoethyl Aspartamide (California, USA, Dionex)) separated subsequently carried on a nano-LC C18 column (200 Å, 2 µm, 75 µm × 25 cm). The nano-LC is coupled online to a QExactive™-Plus Orbitrap(ThermoScientific) mass spectrometer (MS). The nano-LC eluents were infused to the Orbitrap mass- spectrometer with a capillary at 1.7 KV on a nano-ESI source at a flow rate of 300 nl/min. Data dependent acquisition in positive ion mode

was performed for a selected mass range of 350-1800 m/z at the MS1 level with a resolution of 140,000 and at the MS2 level with a resolution of 17,500. The raw data were analyzed by Proteome Discoverer 2.1 Software (ThermoScientific) using Sequest HT as the search engine against the human UniProt/SwissProt database. The threshold of confidence was set above 99% ensuring a false discovery rate of less than 1%. The list of identified proteins, containing iTRAQ ratios of expression levels over control samples, was generated. Proteome Discoverer 2.1 employs a global analytical methodology.

Statistical analyses

For demographics, group mean differences were calculated by ANOVA with Tukey honest significant differences (HSD) post-hoc comparisons for numerical data and with Fisher exact test for categorical variables (gender, smoking status). All analyses were performed using JMP® version 13 (SAS, Cary, North Carolina 27513, USA).

Only data of those proteins that were detected in all HC and patient samples were used. Using GraphPad Prism analysis of the distribution scope of the acquired MS data we only allowed datasets to transition to further levels of analysis if we found the mass level data to be normally distributed. All raw iTRAQ ratios (condition:Con) were log2 transformed before calculating the mean abundance ratio per experimental group. Subsequently, the fold change expression ratios of log2 transformed mean abundances of patient samples over HC on the one hand, and MDD over BD-D and BD-M over SZ were calculated. Finally, only proteins were withheld as biomarker candidates that showed fold change expression ratios of >2 standard deviations to the patient/HC or patient/patient abundance mean.

Our proteomic analysis pipelines and methodologies have previously been demonstrated to be both rigorous and able to generate highly effective and significant data outputs that give actionable high-dimensionality data appreciation [17–24].

Results

Demographics

After screening, 5 MDD patients, 3 BD-D patients, 4 BD-M patients and 4 SZ patients fulfilled all inclusion criteria. In parallel, 6 age and gender matched healthy controls were included. No significant between-group differences were found for age, gender, BMI and smoking status (Fisher exact, Tukey Honest significant

difference). All patients except for 1 medication free MDD patient used some type of psychopharmacotherapeutics. Six patients (n = 3 SZ, n = 1 BD-M, n = 1BDD, n = 1 MDD) were on monotherapy, while all other patients (72%) took at least 2 different psychopharmacological drugs. For an overview of medication status per disorder, see supplementary table 1.

	MDD	BD-D	BD-M	Sz
Antidepressants				
SSRI	2	1	1	
SNRI	2			
MAOI				
TCA	1			
Other AD	3			
Mood stabilizers (Lithium, valproate)		2	3	
Antipsychotics	2	1	4	3
Benzodiazepines		1	2	2

Supplementary table 1: Medication status of included patients. SSRI = selective serotonin reuptake inhibitor, SSNR= selective serotonin & norepinephrine reuptake inhibitor, TCA = tricyclic antidepressant, MAOI = monoamine oxidase inhibitor.

Compared to controls, all patients showed more depressive symptoms as measured by the Hamilton depression rating scale (HDRS). Among patient groups, BD-D and MDD had significantly higher HDRS scores than BD- M patients. As expected, psychotic symptoms as quantified by the Positive and Negative Syndrome Scale (PANSS) were only detected in SZ patients. SZ patients further scored highest on the Young mania rating scale (YMRS) for manic symptoms, followed by BD-M patients who in their turn differed significantly from bipolar depressed patients. No manic symptoms were demonstrated in MDD patients or HC. Demographic variables and symptom scores are presented in table 1.

	HC	MDD	BD-D	BD-M	SZ
n	6	5	3	4	4
Age	32 ± 8,56	28,6 ± 10,95	24,67 ± 3,06	36,5 ± 11,5	34,5 ± 11,5
Gender (M/F)	2/4	1/4	1/2	1/3	1/3
BMI	25,9 ± 3,01	24,74 ± 4,4	24,67 ± 2,84	31,42 ± 2,8	25,68 ± 6,14
Smoking	3	1	2	2	3
HDRS	1 ± 1,55	20,2 ± 1,79*	21 ± 3,46*	10 ± 6,68	14 ± 4,32*
PANSS-P	7 ± 0	7 ± 0	7 ± 0	7 ± 0	20,25 ± 4,03*

YMRS	0,17 ± 0,41	1,6 ± 1,52	2,67 ± 2,31	13,5 ± 0,58*	18,25 ± 1,5*

Table 1: Demographics of patients with major depressive disorder (MDD), bipolar depression (BD-D), mania (BD-M) and schizophrenia (SZ) and healthy controls (HC). Data are presented as mean ± SD; * $p < 0.05$; significantly different from healthy controls.

Protein retrieval capacity and differential expression from controls

To determine potential diagnostic discriminators for patients with either MDD, BD-D, BD-M or SZ, the proteomes of patients' peripheral blood mononuclear cells (PBMC) were mapped and compared with those of age and gender matched healthy controls.

Quantitative proteomic analysis led to the total detection of 4271 individual proteins in HC. Of these, about 90% were also found in the 4 psychiatric traits (PT). To ensure identification of biomarkers with highest certainty of retrieval, we subsequently quantified the number of HC-detected proteins that were only present in all individuals for each experimental group and found a similar detection rate of ~70% for all experimental groups and HC. As pathological biomarkers are preferably distinguishing between the PT and HC, we last calculated the number of proteins detected in all patients per PT that were differentially expressed from controls. Differentially expressed proteins (DEP) were defined as proteins showing a >2 standard deviation (SD) fold change in expression from controls and on average ~5% DEP were found for all PT. Specifically, BD-D had the highest number of DEP (167), followed by MDD (132 DEP), BD-M (125 DEP) and SZ (102 DEP) (Table 2).

	HC n = 6	**MDD n = 5**	**BD-D n = 3**	**SZ n = 4**	**BD-M n = 4**
Total number of proteins detected	4271 (100%)	3954 (93%)°	3918 (92%)°	3918 (92%)°	3954 (93%)°
Total number of proteins detected in all patients	3197 (75%)	2780 (70%)	2806 (72%)	2806 (72%)	2651 (67%)
DEP vs. HC (+/-2 2SD fold change from HC)	*NA*	132 (5%)	167 (6%)	102 (4%)	125 (5%)

Table 2: Protein retrieval capacity and differential expression per condition. HC: healthy controls; MDD: major depressive disorder; BD-D: bipolar disorder, depressed state; BD-M: bipolar disorder, manic state. °: only proteins that were also detected in HC.

Diagnostic biomarkers discriminating unipolar from bipolar depression

As clinical presentation of depressive symptoms hardly allows to discern the underlying pathology to be either

MDD or BD, as such rendering adequate therapeutic intervention difficult, objective discriminative biomarkers are urgently needed. Therefore, we investigated which proteins showed both the greatest differential expression in patients with depressive symptoms vs. HC and in MDD vs. BD-D. Again, proteins with a fold change expression of >2SD were considered as substantially differentially expressed. Of the proteins showing differential expression from HC in either depressive condition, 66 reached the 2SD significance threshold for differential expression between MDD and BD-D and are as such potential discriminatory biomarkers for MDD vs. BD-D. Only 1 protein was found to be significantly contraregulated in MDD vs. BD-D. Interestingly, this concerns the protein HLA class II histocompatibility antigen, DRB1-16 beta chain (HLA-DRB1), a protein from the HLA-II class, which had been extensively implicated in several psychiatric disorders. HLA-DRB1 expression was significantly (i.e. >2 SD expression fold change from HC) downregulated in MDD (-0.21 fold change in expression from HC) while being significantly upregulated in BD-D (0.38 fold change expression from HC). Moreover, of the 65 additional discriminatory biomarkers, 7 more proteins were HLA molecules, confirming them to be of major interest with regard to the physiopathology of depressive disorders.

A list of all 66 discriminatory biomarkers for MDD and BD-D can be found in table 3.

Protein	**MDD/BDD** 2SD = 0.16	**MDD/HC** 2SD = 0.19	**BDD/HC** 2SD = 0.21
HLA class I histocompatibility antigen, A-24 alpha chain	0.64	0.11	-0.53
Keratin, type II cytoskeletal 1	0.63	-0.01	-0.64
HLA class I histocompatibility antigen, B-18 alpha chain	0.62	0.14	-0.48
HLA class II histocompatibility antigen, DRB1-16 beta chain	-0.59	-0.21	0.38
Keratin, type I cytoskeletal 9	0.58	-0.06	-0.63
HLA class I histocompatibility antigen, A-2 alpha chain	0.47	-0.02	-0.49
Histone H2A type 1	0.46	-0.62	-1.09
Interferon-induced GTP-binding protein Mx1	-0.45	-0.05	0.40
Keratin, type I cytoskeletal 10	0.43	0.04	-0.39
Plexin-A4	-0.39	-0.38	0.01
Retinoblastoma-like protein 1	-0.39	0.19	0.58
High mobility group nucleosome-binding domain-containing protein 4	0.37	0.13	-0.25
HD domain-containing protein 2	0.37	-0.04	-0.41
cGMP-inhibited 3',5'-cyclic phosphodiesterase B	-0.35	-0.42	-0.07
Galectin-10	0.33	0.41	0.08
Beta-hexosaminidase	0.32	-0.33	-0.65
Hemoglobin subunit gamma-1	0.30	0.64	0.33
HLA class I histocompatibility antigen, A-68 alpha chain	-0.30	-0.02	0.28

Non-histone chromosomal protein HMG-14	0.30	0.20	-0.09
HLA class I histocompatibility antigen, Cw-1 alpha chain	0.29	0.30	0.00
Bardet-Biedl syndrome 12 protein	-0.28	-0.33	-0.06
NCK-interacting protein with SH3 domain	-0.27	-0.21	0.06
CAP-Gly domain-containing linker protein 2	-0.26	-0.22	0.05
Keratin, type II cytoskeletal 2 epidermal	0.26	0.00	-0.25
40S ribosomal protein S28	0.25	0.29	0.04
Histone H2A type 2-C	0.25	0.04	-0.21
Neutrophil gelatinase-associated lipocalin	0.25	0.53	0.27
OCIA domain-containing protein 2	0.25	-0.22	-0.47
Mitogen-activated protein kinase 13	0.24	0.02	-0.22
Tubulin beta-3 chain	-0.24	-0.60	-0.36
Histone H1.4	0.24	0.29	0.04
Trem-like transcript 1 protein	-0.24	-0.71	-0.47
Hemoglobin subunit gamma-2	0.23	0.36	0.13
PHD finger protein 6	0.23	-0.10	-0.33
Protein IWS1 homolog	0.23	-0.15	-0.38
Tropomyosin beta chain	-0.23	0.00	0.22
Peptidyl-prolyl cis-trans isomerase G	0.22	-0.09	-0.32
39S ribosomal protein L28, mitochondrial	0.22	-0.01	-0.23
Carcinoembryonic antigen-related cell adhesion molecule 8	0.22	0.27	0.05
Alpha-ketoglutarate-dependent dioxygenase FTO	0.22	-0.04	-0.26
Bactericidal permeability-increasing protein	0.22	0.52	0.30
Endoplasmic reticulum aminopeptidase 2	-0.21	-0.45	-0.24
Lactotransferrin	0.21	0.60	0.39
RNA-binding protein 42	0.21	0.00	-0.21
N-sulphoglucosamine sulphohydrolase	0.21	-0.11	-0.31
Platelet glycoprotein Ib alpha chain	-0.21	-0.23	-0.03
Cytochrome b-c1 complex subunit 9	0.20	-0.26	-0.46
ADP-ribosylation factor-like protein 8B	0.20	0.35	0.15
Platelet glycoprotein VI	-0.20	-0.24	-0.04
Cathelicidin antimicrobial peptide	0.20	0.63	0.43
Protein disulfide-isomerase A5	-0.20	-0.24	-0.04
Macrophage migration inhibitory factor	0.20	-0.11	-0.31
HLA class II histocompatibility antigen, DRB1-11 beta chain	-0.19	-0.46	-0.26
Histone H2B type 3-B	0.19	-0.06	-0.25
Liver carboxylesterase 1	0.19	-0.12	-0.31
E3 ubiquitin-protein ligase RNF123	0.19	-0.06	-0.25
HLA class I histocompatibility antigen, B-7 alpha chain	0.19	-0.08	-0.27
Histone H1.2	0.18	0.20	0.02
GTPase IMAP family member 5	-0.18	-0.26	-0.08
Prenylcysteine oxidase-like	-0.18	0.13	0.31
WD repeat-containing protein 43	0.18	-0.24	-0.42
Mitochondrial-processing peptidase subunit alpha	0.17	-0.12	-0.29

Granzyme H	0.17	-0.35	-0.52
Tyrosine-protein phosphatase non-receptor type 7 (Fragment)	0.17	-0.08	-0.25
Protein preY, mitochondrial	0.17	-0.05	-0.21
Granulysin (Fragment)	0.17	-0.34	-0.51

Table 3: Hierarchically ranked expression ratios of proteins that are differentially expressed between MDD and BD-D on the one hand and between either MDD and HC or BD-D and HC on the other. Proteins are hierarchically ranked according to the largest discriminatory potential between MDD and BD-D. Significant differential expression was defined as > 2SD expression fold change; red: increased expression in pathology vs. HC; green: decreased expression vs. HC; greyed out: non-significant differential expression from HC.

Diagnostic biomarkers discriminating manic from psychotic patients

Also manic and psychotic symptom presentations are not always readily discernible based on clinical phenotype and would benefit from objective biological categorization. Therefore, PBMC proteomes of patients with either manic symptoms within the spectrum of bipolar disorder or psychotic symptoms within the spectrum of schizophrenia were compared to identify symptom-specific discriminatory biomarkers. Of the 2651 proteins retrieved in all BD-M and SZ patients, 72 proved to be both significantly differentially expressed between BD-M and SZ on the one hand and between either of the pathologic conditions and HC on the other. Again, one protein (apolipoprotein C-III, APOC3) was found to be significantly contraregulated in BD-M vs. SZ. APOC3 is involved in triglyceride homeostasis and was downregulated in BD-M (-0.24 fold change in expression from HC) and upregulated in SZ (0.23 fold change in expression from HC). Three other proteins were however more strongly distinctiving BD-M from SZ, irrespective of their expression in control individuals. While APOC3 showed a downregulation of -0.47 in BD-M over SZ, these fold changes were stronger in hemoglobin subunit gamma-1 (-0.72 fold change expression BD-M / SZ), galectin-10 (-0.70 fold change expression BD-M / SZ) and HLA class II histocompatibility antigen, DRB1-11 beta chain (-0.54 fold change expression BD-M / SZ).

A list of all 72 discriminatory biomarkers for BD-M and SZ can be found in table 4.

Protein	**BD-M/SZ** 2SD = 0.16	**BD-M/HC** 2SD = 0.20	**SZ/HC** 2SD = 0.21

Hemoglobin subunit gamma-1	-0.72	0.20	0.91
Galectin-10	-0.70	-0.05	0.64
HLA class I histocompatibility antigen, B-7 alpha chain	0.67	0.16	-0.51
HLA class II histocompatibility antigen, DRB1-11 beta chain	-0.54	-0.79	-0.26
HLA class I histocompatibility antigen, A-34 alpha chain	0.51	0.32	-0.19
Apolipoprotein C-III	-0.47	-0.24	0.23
Alanine--tRNA ligase, mitochondrial	0.45	0.47	0.02
Neutrophil gelatinase-associated lipocalin	-0.45	0.44	0.89
Hemoglobin subunit gamma-2	-0.43	0.11	0.54
Retinoblastoma-like protein 1	-0.43	0.38	0.81
Matrix metalloproteinase-9	-0.41	0.39	0.80
Protein S100-A12	-0.41	0.22	0.62
Cathelicidin antimicrobial peptide	-0.40	0.49	0.89
Bardet-Biedl syndrome 12 protein	-0.38	-0.23	0.14
Selenium-binding protein 1	-0.35	0.03	0.38
Granulysin (Fragment)	-0.34	-0.54	-0.19
Zinc finger ZZ-type and EF-hand domain-containing protein 1	-0.34	-0.05	0.29
Non-histone chromosomal protein HMG-17	-0.32	-0.05	0.27
AP-3 complex subunit sigma-1	-0.31	-0.02	0.29
Tubulin beta-3 chain	-0.30	-0.57	-0.27
Hemoglobin subunit alpha	-0.30	0.21	0.51
40S ribosomal protein S21	0.30	0.23	-0.06
Neutrophil defensin 1	0.29	1.02	0.73
Hemoglobin subunit beta	-0.29	0.28	0.57
HLA class I histocompatibility antigen, B-18 alpha chain	-0.29	-0.32	-0.03
Apolipoprotein C-I	-0.29	-0.44	-0.15
Protein Mpv17	-0.29	-0.57	-0.28
Gamma-tubulin complex component 2	-0.28	-0.07	0.21
Lactotransferrin	-0.28	0.51	0.79
C-C motif chemokine 5	0.28	0.26	-0.02
Histone H2A type 2-C	-0.28	-0.22	0.06
Calumenin	0.28	0.22	-0.06
cGMP-inhibited 3',5'-cyclic phosphodiesterase B	-0.27	-0.39	-0.12
Rap guanine nucleotide exchange factor 3	-0.27	-0.27	-0.01
Phosphomevalonate kinase	-0.26	-0.04	0.22
40S ribosomal protein S28	0.26	0.26	-0.01
Zinc finger protein 648	-0.26	0.26	0.52
Bactericidal permeability-increasing protein	-0.26	0.43	0.68
Phospholipase D3	-0.25	-0.23	0.03
Myristoylated alanine-rich C-kinase substrate	0.23	0.30	0.07
Nicotinamide phosphoribosyltransferase	-0.23	0.00	0.22
Band 3 anion transport protein	-0.22	0.04	0.26
CAP-Gly domain-containing linker protein 2	-0.22	-0.26	-0.04

Carcinoembryonic antigen-related cell adhesion molecule 8	-0.22	0.32	0.54
Putative beta-actin-like protein 3	-0.22	-1.07	-0.86
HLA class I histocompatibility antigen, A-2 alpha chain	-0.22	-0.32	-0.10
HLA class I histocompatibility antigen, A-24 alpha chain	-0.21	-0.36	-0.15
Pre-mRNA-splicing factor SYF1	0.21	0.23	0.03
Protein S100-P	-0.21	0.32	0.52
HLA class I histocompatibility antigen, alpha chain E	-0.20	-0.40	-0.20
Granzyme H	0.20	-0.18	-0.38
Condensin complex subunit 1	-0.20	-0.31	-0.11
Cytochrome b-c1 complex subunit 9	-0.20	-0.34	-0.15
Keratin, type I cytoskeletal 9	-0.20	-0.29	-0.09
Ubiquitin-conjugating enzyme E2 D1	0.19	0.25	0.05
Small nuclear ribonucleoprotein E	-0.19	-0.22	-0.03
Ubiquinone biosynthesis monooxygenase COQ6, mitochondrial	-0.19	-0.37	-0.18
Apolipoprotein A-I	-0.18	-0.29	-0.11
Carbonic anhydrase 1	-0.18	0.03	0.22
Lysosomal alpha-glucosidase	-0.18	-0.35	-0.17
Haptoglobin	-0.18	0.04	0.22
COMM domain-containing protein 4	-0.18	-0.29	-0.11
Eukaryotic elongation factor 2 kinase	-0.17	-0.61	-0.44
Mitochondrial carrier homolog 1 (Fragment)	-0.17	-0.39	-0.21
Protein S100-A8	-0.17	0.53	0.70
Peroxiredoxin-2	-0.17	0.05	0.22
E3 ubiquitin-protein ligase TRIM22	-0.17	-0.26	-0.09
Tubulin alpha-8 chain (Fragment)	0.17	-0.09	-0.26
Macrophage migration inhibitory factor	0.17	-0.06	-0.22
H(+)/Cl(-) exchange transporter 3	0.17	0.33	0.17
Cathepsin G	0.16	0.55	0.38
Solute carrier family 2, facilitated glucose transporter member 1	-0.16	0.12	0.28

Table 4: Expression ratios of proteins that are differentially expressed between BD-M and SZ on the one hand and between either BD-M and HC or SZ and HC on the other. Proteins are hierarchically ranked according to the largest discriminatory potential between BD-M and SZ. Significant differential expression was defined as > 2SD expression fold change; red: increased expression in pathology vs. HC; green: decreased expression vs. HC; greyed out: non-significant differential expression from HC.

Discussion

With this work, we aimed at establishing a list of potential biomarkers that show great promise in objectively categorizing patients with MDD vs. BD-D on the one hand, and BD-M vs. SZ on the other. Non-targeted LCMS-

based proteomics of patient PBMC revealed 66 proteins that may enable biological tracking of depressive symptoms to either a unipolar or bipolar depression context. Likewise, 72 proteins might biologically differentiate between manic and psychotic symptoms. For both comparisons, a single protein was found to be significantly contraregulated vs. HC between the 2 pathologies. In case of depression, this concerns the HLA class II histocompatibility antigen, DRB1-16 beta chain. When comparing differential expression vs. HC between BD-M and SZ, only apolipoprotein C-III was found to be significantly contraregulated.

By identifying proteins that both differ between the 2 comparative diseases and between either of the pathologies and healthy individuals, our data allow for 2 methods of clinical biomarking. The first method would be to utilize proteins that show contraregulation or differential regulation in either of the pathologies vs. HC. With this approach, a reference range of these proteins should be determined in a large cohort of HC that can subsequently be used as threshold value for patient samples. Preferably, we suggest the potential HC-based biomarker proteins to either show contraregulation vs. HC (i.e. HLA-DRB1-16b for MDD vs. BD-D and APOC3 for BD-M vs. SZ) or show less than 1 SD fold change differential expression in 1 of the 2 comparative pathologies while being significantly dysregulated in the other. Specifically, for MDD vs. BD-D, 35 proteins show < 1 SD fold change expression in either of the 2 pathologies and > 2SD fold change expression in the other and could thus be further validated as HC-based biomarkers for discriminatory diagnosis of MDD and BD-D (see Supplementary table 2).

Biomarker MDD vs. HC		**Biomarker BD-D vs. HC**	
Plexin-A4	↓	Keratin, type II cytoskeletal 1	↓
cGMP-inhibited 3',5'-cyclic phosphodiesterase B	↓	Keratin, type I cytoskeletal 9	↓
Galectin-10	↑	HLA class I histocompatibility antigen, A-2 alpha chain	↓
Non-histone chromosomal protein HMG-14	↑	Interferon-induced GTP-binding protein Mx1	↑
HLA class I histocompatibility antigen, Cw-1 alpha chain	↑	Keratin, type I cytoskeletal 10	↓
Bardet-Biedl syndrome 12 protein	↓	HD domain-containing protein 2	↓
NCK-interacting protein with SH3 domain	↓	HLA class I histocompatibility antigen, A-68 alpha chain	↑
CAP-Gly domain-containing linker protein 2	↓	Keratin, type II cytoskeletal 2 epidermal	↓
40S ribosomal protein S28	↑	Histone H2A type 2-C	↓
Histone H1.4	↑	Mitogen-activated protein kinase 13	↓
Carcinoembryonic antigen-related cell	↑	Tropomyosin beta chain	↑

adhesion molecule 8			
Platelet glycoprotein Ib alpha chain	↓	Peptidyl-prolyl cis-trans isomerase G	↓
Platelet glycoprotein VI	↓	39S ribosomal protein L28, mitochondrial	↓
Protein disulfide-isomerase A5	↓	Alpha-ketoglutarate-dependent dioxygenase FTO	↓
Histone H1.2	↑	RNA-binding protein 42	↓
GTPase IMAP family member 5	↓	Histone H2B type 3-B	↓
		E3 ubiquitin-protein ligase RNF123	↓
		HLA class I histocompatibility antigen, B-7 alpha chain	↓
		Tyrosine-protein phosphatase non-receptor type 7 (Fragment)	↓
		Protein preY, mitochondrial	↓

Supplementary table 2: Overview of HC-based potential biomarkers for MDD (left) and BD-D (right). ↓: significant downregulation vs. HC; ↑: significant upregulation vs. HC.

With regard to objectively discerning BD-M from SZ, 31 proteins show < 1 SD fold change expression in either of the 2 pathologies and > 2SD fold change expression in the other and could thus be further validated as HC-based biomarkers for discriminatory diagnosis of BD-M and SZ (see Supplementary table 3).

Biomarker BD-M vs. HC		Biomarker SZ vs. HC	
Alanine--tRNA ligase, mitochondrial	↑	Galectin-10	↑
40S ribosomal protein S21	↑	Selenium-binding protein 1	↑
HLA class I histocompatibility antigen, B-18 alpha chain	↓	Zinc finger ZZ-type and EF-hand domain-containing protein 1	↑
C-C motif chemokine 5	↑	Non-histone chromosomal protein HMG-17	↑
Histone H2A type 2-C	↓	AP-3 complex subunit sigma-1	↑
Calumenin	↑	Gamma-tubulin complex component 2	↑
Rap guanine nucleotide exchange factor 3	↓	Phosphomevalonate kinase	↑
40S ribosomal protein S28	↑	Nicotinamide phosphoribosyltransferase	↑
Phospholipase D3	↓	Band 3 anion transport protein	↑
Myristoylated alanine-rich C-kinase substrate	↑	Carbonic anhydrase 1	↑
CAP-Gly domain-containing linker protein 2	↓	Haptoglobin	↑
HLA class I histocompatibility antigen, A-2 alpha chain	↓	Peroxiredoxin-2	↑
Pre-mRNA-splicing factor SYF1	↑	Tubulin alpha-8 chain (Fragment)	↓
Keratin, type I cytoskeletal 9	↓	Macrophage migration inhibitory factor	↓
Ubiquitin-conjugating enzyme E2 D1	↑		
Small nuclear ribonucleoprotein E	↓		
E3 ubiquitin-protein ligase TRIM22	↓		

Supplementary table 3: Overview of HC-based potential biomarkers for BD-M (left) and SZ (right). ↓: significant downregulation vs. HC; ↑: significant upregulation vs. HC.

An alternative strategy to clinically implement proteins retrieved in this study as discriminatory biomarkers is to determine reference ranges for a large cohort of patients from both comparative pathologies in order to be able to quantifiably categorize patients to either of the 2 diagnostic groups. As this approach would per definition require biomarkers that show the largest discrepancy between the 2 comparative pathologies, the proteins ranked highest in the respective MDD / BD-D and BD-M / SZ biomarker lists (see tables 3 and 4) would theoretically show the largest predictive potential. Although considerably more time- and effort consuming, this approach would undoubtedly prove more sensitive and specific than less robust but considerably faster HC-based biomarker optimization.

Both approaches however, require thorough follow-up research in order to affirm true biomarker capacity of the candidate proteins identified within this project. In a first step, reproducibility of the results should be substantiated in a larger training cohort of patients as this project concerned a proof of concept with a limited sample size. Subsequently, retrieval capacity of the confirmed biomarker candidates by more efficient and accessible antigen-based methods like single or multiplex enzyme-linked immuno-assays (ELISA) should be evaluated as this would be the technology of choice for low-threshold implementation in clinic. For those candidate biomarkers retrievable as such, reference ranges should then be established for either a large HC or both patient populations and lastly, the predictive capacity of the final selection of candidate biomarkers is to be determined by comparing the accuracy rate of blind objective diagnostication of a large test sample of patients based on biomarker quantification with the phenotypical diagnosis made by the treating psychiatrist. If none of the above individual candidate biomarkers prove valid in having sufficient predictive capacity, machine learning algorithms might be applied to these data to distill a molecular fingerprint composed of differential regulation of multiple markers.

Although this cross-sectional study has merit in aiming to objectively discriminate between the aforementioned diagnoses to determine adequate treatment regimens, it does not provide information on biomarker stability throughout time. Longitudinal follow-up studies could provide

insight in biomarker fluctuation according to disease course and moreover, potentially reveal biomarkers correlating to either disease state or treatment effectiveness. Of note, correlation analyses on our data revealed no correlation between the 5 most discriminating MDD vs. BD-D proteins and HDRS on the one hand, and between the 5 most discriminating BD-M vs. SZ proteins and either the YMRS or the PANSS-P clinical scales on the other (data not shown). This is to be expected as both BD-D and MDD patients would score high on the HDRS and likewise, BD-M and SZ patients would have some overlap in YMRS and PANSS-P scoring.

As the aim of this study is to find biomarkers that are readily accessible, peripheral blood cells were chosen as investigation medium. This implies that at the current study stage we do not have any information on the biomarker's distribution in the brain. As some biomarkers might be subject to active transport across the blood- brain barrier following excretion from PBMC, they might additionally function as disease-related state or trait biomarkers in the brain. We are however hesitant to draw any definite pathophysiological conclusions from our data as this project concerns intracellular protein detection in peripherally circulating blood cells and may therefore be of little representative value for central mechanistic neurobiological processes. Notwithstanding, several of our biomarker candidates have already been implicated in psychiatric afflictions. With 16 out of the total of 138 (11%) potential biomarkers for MDD vs. BD-D and BD-M vs. SZ, the human leukocyte antigen (HLA) family is highly represented in our candidate list. HLA proteins are cell-surface proteins originating from the major histocompatibility (MHC) gene locus that regulate the body's immune responses and are present on all bodily cells except red blood cells. Notably, a recent study has indicated that the strongest genetic association for the risk of schizophrenia development is linked to greater expression of the *complement component 4 (C4) A* allele, contained on the MHC locus 25. Although associations between certain HLA haplotypes and mood disorders and schizophrenia have been extensively described 26–30, the potential of HLA peptides or proteins as predictive diagnostic biomarkers has to our knowledge never been explored. In parallel, a considerable number of diagnostic protein candidates are involved in immune processes. While

this is not unexpected considering the immune mediating function of PBMC, it is in line with the recently re-emerging hypotheses on the role of increased inflammatory processes in mediating psychiatric afflictions (for review see

31–33,34,35).

Some contemplations might be taken into account when considering these data. First, as mentioned above, this project concerns a proof of concept study regarding the feasibility of proteome-based diagnostics in objectively discriminating MDD from BD-D and BD-M from SZ. While our small data cohort was able to generate distinctive and actionable data streams it is clear that further more advanced studies with considerably bigger patient cohorts will be more desirable for predictive diagnostic activities. These studies are currently underway in our laboratory.

Additionally, patients in this study had been recently started on medication, which might in itself influence

protein expression. Replication in drug-free patients and investigation of the potential influence of different types of psychopharmacological treatment regimens on expression of PBMC biomarkers in larger cohorts are envisaged. Lastly, we only considered proteins as candidate biomarkers if they were not only strongly differentially expressed between the 2 comparative pathologies, but were in addition also significantly dysregulated when compared to healthy individuals. In theory, it could be expected that some proteins with high pathological discriminative capacity would fall within the 2 SD fold change expression requirement with HC and are thus not registered in our final biomarker candidate bank.

Nonetheless, we believe that this work provides a major stepping stone towards imminent implementation of objective biological discriminatory diagnostics in the field of clinical psychiatry.

Acknowledgements

The authors thank study nurses Eline Schellens and Ingrid Vanderplas and psychological assistant Anneke Coeck for their extensive work in performing clinical trial interventions. The study was

approved by the ethics committee of the University Hospital Antwerp, Belgium.

Disclosure Statement

The authors have no conflict of interest to declare.

Funding Sources

The study was performed thanks to the kind support of the Psychiatry Hospital of Duffel, Belgium, who provided consumable costs and personnel funding. Additional funding was provided by a research grant of the University of Antwerp, Belgium.

Author Contributions

VC performed biomarker candidate analyses and wrote the paper. ODW, JG, SM and JH wrote the paper. ODW, JG, JH and SM performed LCMS analyses. JVG and AA performed experimental LCMS procedures. TL and AVS performed clinical testing and demographic analyses. MM supervised the project. All authors revised the manuscript.

References

1. Angst, J. *et al.* Prevalence and characteristics of undiagnosed bipolar disorders in patients with a major depressive episode: the BRIDGE study. *Arch. Gen. Psychiatry* **68**, 791–798 (2011).
2. Hirschfeld, R. M. A., Lewis, L. & Vornik, L. A. Perceptions and impact of bipolar disorder: how far have we really come? Results of the national depressive and manic-depressive association 2000 survey of individuals with bipolar disorder. *J. Clin. Psychiatry* **64**, 161–174 (2003).
3. Pini, S. *et al.* Cross-sectional similarities and differences between schizophrenia, schizoaffective disorder and mania or mixed mania with mood-incongruent psychotic features. *Eur. Psychiatry* **19**, 8– 14 (2004).

4. Benabarre, A. *et al.* Bipolar disorder, schizoaffective disorder and schizophrenia: epidemiologic, clinical and prognostic differences. *Eur. Psychiatry* **16**, 167–172 (2001).

5. Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R. & Walters, E. E. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry* **62**, 617–627 (2005).

6. Liu, B. *et al.* Efficacy and safety of long-term antidepressant treatment for bipolar disorders - A meta- analysis of randomized controlled trials. *J. Affect. Disord.* **223**, 41–48 (2017).

7. Goodwin, G. M. *et al.* Evidence-based guidelines for treating bipolar disorder: Revised third edition recommendations from the British Association for Psychopharmacology. *J. Psychopharmacol.* **30**, 495– 553 (2016).

8. Viktorin, A. *et al.* The risk of switch to mania in patients with bipolar disorder during treatment with an antidepressant alone and in combination with a mood stabilizer. *Am. J. Psychiatry* **171**, 1067–1073 (2014).

9. Altamura, A. C. *et al.* Misdiagnosis, duration of untreated illness (DUI) and outcome in bipolar patients with psychotic symptoms: A naturalistic study. *J. Affect. Disord.* **182**, 70–75 (2015).

10. Cipriani, A. *et al.* Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple- treatments meta-analysis. *Lancet* **378**, 1306–1315 (2011).

11. Altamura, A. C. & Goikolea, J. M. Differential diagnoses and management strategies in patients with

schizophrenia and bipolar disorder. *Neuropsychiatr. Dis. Treat.* **4**, 311–317 (2008).

12. Chakrabarty, T. & Yatham, L. N. Objective and biological markers in bipolar spectrum presentations.

Expert Rev. Neurother. **19**, 195–209 (2019).

13. Pelsers, M. M. *et al.* Influence of age and sex and day-to-day and within-day biological

variation on plasma concentrations of fatty acid-binding protein and myoglobin in healthy subjects. *Clin. Chem.* **45**, 441–443 (1999).

14. Macy, E. M., Hayes, T. E. & Tracy, R. P. Variability in the measurement of C-reactive protein in healthy subjects: implications for reference intervals and epidemiological applications. *Clin. Chem.* **43**, 52–58 (1997).
15. Garde, A. H., Hansen, A. M., Skovgaard, L. T. & Christensen, J. M. Seasonal and biological variation of blood concentrations of total cholesterol, dehydroepiandrosterone sulfate, hemoglobin A(1c), IgA, prolactin, and free testosterone in healthy women. *Clin. Chem.* **46**, 551–559 (2000).
16. Johansen, J. S. *et al.* Diurnal, weekly, and long-time variation in serum concentrations of YKL-40 in healthy subjects. *Cancer Epidemiol. Biomarkers Prev.* **17**, 2603–2608 (2008).
17. Kashyap, M. K. *et al.* SILAC-based quantitative proteomic approach to identify potential biomarkers from the esophageal squamous cell carcinoma secretome. *Cancer Biol. Ther.* **10**, 796–810 (2010).
18. Park, S.-S. *et al.* Effective correction of experimental errors in quantitative proteomics using stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteomics* **75**, 3720–3732 (2012).
19. Park, S.-S. & Maudsley, S. Discontinuous pH gradient-mediated separation of TiO2-enriched phosphopeptides. *Anal. Biochem.* **409**, 81–88 (2011).
20. Wu, X. *et al.* Quantitative phosphoproteomic analysis reveals reciprocal activation of receptor tyrosine kinases between cancer epithelial cells and stromal fibroblasts. *Clin. Proteomics* **15**, 21 (2018).
21. Mattison, J. A. *et al.* Resveratrol prevents high fat/sucrose diet-induced central arterial wall inflammation and stiffening in nonhuman primates. *Cell Metab.* **20**, 183–190 (2014).
22. Lu, D. *et al.* Nuclear GIT2 is an ATM substrate and promotes DNA repair. *Mol. Cell. Biol.*

35, 1081– 1096 (2015).

23. Wu, W. W., Shen, R.-F., Park, S.-S., Martin, B. & Maudsley, S. Precursor ion exclusion for enhanced identification of plasma biomarkers. *Proteomics Clin. Appl.* **6**, 304–308 (2012).
24. Cutler, J. A. *et al.* Integrative phosphoproteome and interactome analysis of the role of Ubash3b in BCR-ABL signaling. *Leukemia* **34**, 301–305 (2020).
25. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).
26. Stancer, H. C. *et al.* Confirmation of the relationship of HLA (chromosome 6) genes to depression and manic depression. II. The Ontario follow-up and analysis of 117 kindreds. *Ann. Hum. Genet.* **52**, 279– 298 (1988).
27. Debnath, M. *et al.* The HLA-G low expressor genotype is associated with protection against bipolar disorder. *Hum. Immunol.* **74**, 593–597 (2013).
28. Sundaresh, A. *et al.* The HLA-G Genetic Contribution to Bipolar Disorder: A Trans-Ethnic Replication. *Immunol. Invest.* **47**, 593–604 (2018).
29. Rajasekaran, A. *et al.* The impact of HLA-G 3' UTR variants and sHLA-G on risk and clinical correlates of schizophrenia. *Hum. Immunol.* **77**, 1166–1171 (2016).
30. Morgan, L. Z. *et al.* Quantitative Trait Locus and Brain Expression of HLA-DPA1 Offers Evidence of Shared Immune Alterations in Psychiatric Disorders. *Microarrays (Basel)* **5**, (2016).
31. Pape, K., Tamouza, R., Leboyer, M. & Zipp, F. Immunoneuropsychiatry - novel perspectives on brain disorders. *Nat. Rev. Neurol.* (2019) doi:10.1038/s41582-019-0174-4.
32. Bauer, M. E. & Teixeira, A. L. Inflammation in psychiatric disorders: what comes first? *Ann. N. Y. Acad. Sci.* **1437**, 57–67 (2019).
33. De Picker, L. J., Morrens, M., Chance, S. A. & Boche, D. Microglia and Brain Plasticity in Acute Psychosis and Schizophrenia Illness Course: A Meta-Review. *Front. Psychiatry* **8**, (2017).

34. De Picker, L. *et al.* State-associated changes in longitudinal [F]-PBR111 TSPO PET Imaging of psychosis patients: evidence for the accelerated ageing hypothesis? *Brain Behav. Immun.* (2018) doi:10.1016/j.bbi.2018.11.318.

35. van den Ameele, S. *et al.* Markers of Inflammation and Monoamine Metabolism Indicate Accelerated Aging in Bipolar Disorder. *Front. Psychiatry* **9**, 250 (2018).

4. Significantly Enriched Gene Sets

Table S2 Significantly enriched gene sets when expression values of MDD and BD-D patients are compared.

Gene set	Direction	P-value	Q-value
REACTOME_RRNA_PROCESSING	Up	3,2993E-07	1,5870E-04
GNF2_DAP3	Up	1,5379E-05	3,3564E-03
GNF2_FBL	Up	2,1436E-05	4,3028E-03
REACTOME_FORMATION_OF_FIBRIN_CLOT_CLOTTING_CASCADE	Down	4,2025E-05	7,1262E-03
GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS	Up	7,8034E-05	1,0819E-02
GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY	Up	1,0155E-04	1,2532E-02
GO_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE	Up	1,2837E-04	1,4905E-02
GO_STRUCTURAL_CONSTITUENT_OF_RIBOSOME	Up	1,6969E-04	1,7619E-02
GO_CYTOSOLIC_RIBOSOME	Up	1,7831E-04	1,7943E-02
GO_RIBOSOMAL_SUBUNIT	Up	1,8956E-04	1,8576E-02
KEGG_RIBOSOME	Up	2,5244E-04	2,2594E-02
REACTOME_NONSENSE_MEDIATED_DECAY_NMD	Up	2,5946E-04	2,2671E-02
GO_RIBOSOME	Up	3,6075E-04	2,6516E-02
MODULE_32	Up	3,7526E-04	2,7309E-02
REACTOME_SELENOAMINO_ACID_METABOLISM	Up	4,8309E-04	3,2471E-02
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY	Up	4,8993E-04	3,2471E-02
GO_RIBOSOMAL_LARGE_SUBUNIT_BIOGENESIS	Up	5,0161E-04	3,2572E-02
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION	Up	5,3237E-04	3,3822E-02
GSE6269_HEALTHY_VS_STAPH_PNEUMO_INF_PBMC_DN	Down	5,5766E-04	3,4590E-02
GNF2_ST13	Up	7,6828E-04	4,1566E-02
REACTOME_REGULATION_OF_EXPRESSION_OF_SLITS_AND_ROBOS	Up	7,6888E-04	4,1566E-02
REACTOME_COMMON_PATHWAY_OF_FIBRIN_CLOT_FORMATION	Down	8,2857E-04	4,3107E-02
GO_PLATELET_ACTIVATION	Down	6,8600E-15	1,6169E-10
GO_PLATELET_AGGREGATION	Down	1,6805E-14	1,9804E-10
GO_HOMOTYPIC_CELL_CELL_ADHESION	Down	1,3729E-13	1,0786E-09
REACTOME_PLATELET_ACTIVATION_SIGNALING_AND_AGGREGATION	Down	1,0779E-12	6,3517E-09
GO_PLATELET_DEGRANULATION	Down	2,4127E-12	1,1373E-08
REACTOME_RESPONSE_TO_ELEVATED_PLATELET_CYTOSOLIC_CA2	Down	8,4149E-11	3,3057E-07
WIERENGA_STAT5A_TARGETS_DN	Down	1,3680E-10	4,6061E-07
GO_RNA_PROCESSING	Up	1,7100E-10	5,0380E-07
GO_PLATELET_ALPHA_GRANULE	Down	2,7492E-10	7,2000E-07
GO_WOUND_HEALING	Down	4,5939E-10	1,0828E-06
KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM	Down	1,4131E-09	3,0280E-06
GO_RESPONSE_TO_WOUNDING	Down	4,2857E-09	6,3134E-06
REACTOME_PLATELET_ADHESION_TO_EXPOSED_COLLAGEN	Down	4,6810E-09	6,4161E-06
REACTOME_HEMOSTASIS	Down	4,8998E-09	6,4161E-06
GO_MRNA_PROCESSING	Up	9,4458E-09	1,1718E-05
GO_COAGULATION	Down	1,1565E-08	1,2980E-05
REACTOME_PLATELET_AGGREGATION_PLUG_FORMATION	Down	1,4117E-08	1,5125E-05
REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA	Up	1,4912E-08	1,5281E-05
RAGHAVACHARI_PLATELET_SPECIFIC_GENES	Down	1,5908E-08	1,5623E-05
GO_REGULATION_OF_PLATELET_ACTIVATION	Down	1,9876E-08	1,7659E-05

GO_RNA_SPLICING	Up	2,0229E-08	1,7659E-05
TENEDINI_MEGAKARYOCYTE_MARKERS	Down	2,5560E-08	2,0774E-05
HECKER_IFNB1_TARGETS	Down	3,1587E-08	2,4016E-05
KEGG_FOCAL_ADHESION	Down	5,6252E-08	3,8996E-05
GO_REGULATION_OF_HOMOTYPIC_CELL_CELL_ADHESION	Down	7,9136E-08	5,1076E-05
GO_REGULATION_OF_PLATELET_AGGREGATION	Down	9,5881E-08	5,7947E-05
GSE45365_HEALTHY_VS_MCMV_INFECTION_CD11B_DC_DN	Down	1,4737E-07	8,4721E-05
GO_CELL_MATRIX_ADHESION	Down	1,5166E-07	8,5107E-05
KEGG_DILATED_CARDIOMYOPATHY	Down	1,7156E-07	9,4041E-05
KEGG_ECM_RECEPTOR_INTERACTION	Down	2,4680E-07	1,2927E-04
GO_REGULATION_OF_BODY_FLUID_LEVELS	Down	2,5856E-07	1,3248E-04
GO_PLATELET_ALPHA_GRANULE_LUMEN	Down	3,6585E-07	1,7246E-04
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	Down	6,0594E-07	2,6448E-04
GO_CELL_SUBSTRATE_ADHESION	Down	6,3037E-07	2,6842E-04
GSE45365_WT_VS_IFNAR_KO_BCELL_MCMV_INFECTION_DN	Down	6,6424E-07	2,7467E-04
SRF_Q5_01	Down	1,8848E-06	7,0514E-04
GO_ACTOMYOSIN	Down	3,2447E-06	1,0925E-03
GO_REGULATION_OF_EXOCYTOSIS	Down	3,4386E-06	1,1415E-03
GO_ACTIN_FILAMENT_BUNDLE	Down	4,0076E-06	1,2501E-03
GO_SUBSTRATE_ADHESION_DEPENDENT_CELL_SPREADING	Down	4,0826E-06	1,2501E-03
CHEN_LVAD_SUPPORT_OF_FAILING_HEART_UP	Down	4,8863E-06	1,4765E-03
GSE34156_TLR1_TLR2_LIGAND_VS_NOD2_AND_TLR1_TLR2_LIGAND_24H_TREATED_MONOCYTE_UP	Down	5,6186E-06	1,6554E-03
REACTOME_SMOOTH_MUSCLE_CONTRACTION	Down	6,3190E-06	1,8387E-03
GO_STRUCTURAL_CONSTITUENT_OF_MUSCLE	Down	7,6511E-06	2,0296E-03
REACTOME_INTEGRIN_SIGNALING	Down	7,6528E-06	2,0296E-03
SRF_Q4	Down	7,6639E-06	2,0296E-03
GO_REGULATION_OF_WOUND_HEALING	Down	7,9476E-06	2,0361E-03
GO_NEGATIVE_REGULATION_OF_PLATELET_ACTIVATION	Down	9,0763E-06	2,2758E-03
GO_NEGATIVE_REGULATION_OF_PLATELET_AGGREGATION	Down	9,0763E-06	2,2758E-03
GO_RAB_PROTEIN_SIGNAL_TRANSDUCTION	Down	9,6363E-06	2,3659E-03
GO_EXTRACELLULAR_MATRIX_BINDING	Down	1,0090E-05	2,4268E-03
GO_NEGATIVE_REGULATION_OF_HOMOTYPIC_CELL_CELL_ADHESION	Down	1,0720E-05	2,5522E-03
REACTOME_GRB2_SOS_PROVIDES_LINKAGE_TO_MAPK_SIGNALING_FOR_INTEGRINS	Down	1,2351E-05	2,8823E-03
GO_EXOCYTOSIS	Down	1,2750E-05	2,9462E-03
JISON_SICKLE_CELL_DISEASE_UP	Down	1,3061E-05	2,9888E-03
GSE6269_HEALTHY_VS_FLU_INF_PBMC_DN	Down	1,4020E-05	3,1514E-03
GNATENKO_PLATELET_SIGNATURE	Down	1,5887E-05	3,4070E-03
PASINI_SUZ12_TARGETS_DN	Down	1,8740E-05	3,8745E-03
REACTOME_RAB_GERANYLGERANYLATION	Down	1,9188E-05	3,9328E-03
REACTOME_P130CAS_LINKAGE_TO_MAPK_SIGNALING_FOR_INTEGRINS	Down	2,0182E-05	4,1007E-03
GO_REGULATION_OF_RESPONSE_TO_WOUNDING	Down	2,1708E-05	4,3028E-03
GSE34156_NOD2_LIGAND_VS_TLR1_TLR2_LIGAND_6H_TREATED_MONOCYTE_DN	Down	2,1724E-05	4,3028E-03
GO_REGULATION_OF_COAGULATION	Down	2,6462E-05	5,0708E-03
REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION	Down	2,7519E-05	5,2308E-03
SRF_01	Down	2,8081E-05	5,2313E-03

GSE24142_EARLY_THYMIC_PROGENITOR_VS_DN3_THYMOCYTE_UP	Down	2,9824E-05	5,3254E-03
GO_MYOFILAMENT	Down	3,0496E-05	5,4044E-03
RICKMAN_HEAD_AND_NECK_CANCER_F	Down	3,7914E-05	6,5278E-03
GO_SECRETION	Down	3,7943E-05	6,5278E-03
REACTOME_MAP2K_AND_MAPK_ACTIVATION	Down	4,6405E-05	7,5956E-03
REACTOME_TRANSPORT_OF_MATURE_TRANSCRIPT_TO_CYTOPLASM	Up	4,7481E-05	7,6260E-03
GSE29949_CD8_POS_DC_SPLEEN_VS_DC_BRAIN_DN	Down	4,7792E-05	7,6260E-03
GO_GTPASE_ACTIVITY	Down	4,7885E-05	7,6260E-03
GO_MRNA_EXPORT_FROM_NUCLEUS	Up	5,2305E-05	7,9538E-03
GO_POSITIVE_REGULATION_OF_EPITHELIAL_CELL_MIGRATION	Down	5,4870E-05	8,2136E-03
GO_GUANYL_NUCLEOTIDE_BINDING	Down	6,6419E-05	9,7236E-03
GO_PIGMENT_GRANULE	Down	7,0604E-05	1,0147E-02
GO_MUSCLE_FILAMENT_SLIDING	Down	8,1006E-05	1,1166E-02
PID_INTEGRIN1_PATHWAY	Down	8,3585E-05	1,1386E-02
GO_SECRETORY_VESICLE	Down	8,4056E-05	1,1386E-02
MANALO_HYPOXIA_UP	Down	8,8864E-05	1,1641E-02
GSE34156_UNTREATED_VS_24H_NOD2_LIGAND_TREATED_MONOCYTE_DN	Down	9,7214E-05	1,2188E-02
HALLMARK_MYOGENESIS	Down	9,8260E-05	1,2254E-02
REACTOME_STRIATED_MUSCLE_CONTRACTION	Down	1,0545E-04	1,2945E-02
GO_EXTRACELLULAR_STRUCTURE_ORGANIZATION	Down	1,0997E-04	1,3361E-02
PICCALUGA_ANGIOIMMUNOBLASTIC_LYMPHOMA_UP	Down	1,1962E-04	1,4168E-02
GO_SECRETORY_GRANULE	Down	1,4686E-04	1,6483E-02
GO_POSITIVE_REGULATION_OF_EXOCYTOSIS	Down	1,5919E-04	1,7368E-02
GO_GDP_BINDING	Down	1,6220E-04	1,7457E-02
REACTOME_MUSCLE_CONTRACTION	Down	1,6696E-04	1,7490E-02
REACTOME_EPH_EPHRIN_SIGNALING	Down	2,2239E-04	2,0396E-02
GO_PROTEIN_ACTIVATION_CASCADE	Down	2,2841E-04	2,0706E-02
GSE22886_NAIVE_CD8_TCELL_VS_DC_DN	Down	2,5722E-04	2,2671E-02
GO_DNA_TEMPLATED_TRANSCRIPTION_ELONGATION	Up	2,5970E-04	2,2671E-02
REACTOME_SIGNALING_BY_MODERATE_KINASE_ACTIVITY_BRAF_MUTANTS	Down	2,7537E-04	2,3272E-02
GO_FIBRONECTIN_BINDING	Down	2,8319E-04	2,3586E-02
GO_MAST_CELL_ACTIVATION	Down	2,8826E-04	2,3673E-02
GSE22886_NAIVE_CD4_TCELL_VS_DC_DN	Down	2,9300E-04	2,3716E-02
GSE24142_DN2_VS_DN3_THYMOCYTE_FETAL_UP	Down	2,9481E-04	2,3716E-02
GO_MUSCLE_ORGAN_MORPHOGENESIS	Down	2,9859E-04	2,3938E-02
GO_SMOOTH_MUSCLE_CELL_MIGRATION	Down	3,0276E-04	2,4071E-02
MODULE_131	Down	3,0369E-04	2,4071E-02
GO_CARDIAC_MUSCLE_TISSUE_MORPHOGENESIS	Down	3,2078E-04	2,4877E-02
GO_CELL_ACTIVATION	Down	3,5647E-04	2,6369E-02
HSF1_01	Down	4,3158E-04	3,0096E-02
SRF_Q6	Down	4,4267E-04	3,0687E-02
REACTOME_RHO_GTPASES_ACTIVATE_PAKS	Down	4,5418E-04	3,1119E-02
GO_CELL_SURFACE	Down	5,0543E-04	3,2694E-02
GSE10325_CD4_TCELL_VS_BCELL_DN	Down	5,3567E-04	3,3911E-02
GSE3982_NEUTROPHIL_VS_NKCELL_UP	Down	5,5297E-04	3,4480E-02
PID_EPHRINB_REV_PATHWAY	Down	5,9476E-04	3,5945E-02

HALLMARK_COAGULATION	Down	6,1391E-04	3,6819E-02
SMIRNOV_CIRCULATING_ENDOTHELIOCYTES_IN_CANCER_UP	Down	7,3662E-04	4,0555E-02
GSE22886_NAIVE_CD4_TCELL_VS_MONOCYTE_DN	Down	7,7184E-04	4,1630E-02
REACTOME_TRANSPORT_OF_MATURE_MRNAS_DERIVED_FROM_INTRONLESS_TRANSCRIPTS	Up	7,8223E-04	4,1891E-02
GSE24142_DN2_VS_DN3_THYMOCYTE_UP	Down	8,9116E-04	4,4869E-02
GO_PHAGOCYTOSIS	Down	9,9905E-04	4,8155E-02