Universidade de Aveiro

2020

Ana Cristina de Jesus Martinho

IDENTIFICATION OF SCHIZOPHRENIA BIOMARKERS USING A PROTEOMICS APPROACH ON PBMCs

ABORDAGEM PROTEÓMICA PARA IDENTIFICAÇÃO DE BIOMARCADORES EM ESQUIZOFRENIA ATRAVÉS DE PBMCs

Universidade de Aveiro

2020

Ana Cristina de Jesus Martinho

IDENTIFICATION OF SCHIZOPHRENIA BIOMARKERS USING A PROTEOMICS APPROACH ON PBMCs

ABORDAGEM PROTEÓMICA PARA IDENTIFICAÇÃO DE BIOMARCADORES EM ESQUIZOFRENIA ATRAVÉS DE PBMCs

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica do Doutor Bruno Manadas e Doutor João Rodrigues, investigadores do Centro de Neurociências e Biologia Celular (CNC) da Universidade de Coimbra e do Doutor Luís Souto Miranda, Professor Auxiliar do Departamento de Biologia da Universidade de Aveiro.

Este projeto realizado no grupo Life Sciences and Mass Spectrometry do Centro de Neurociências e Biologia Celular (Universidade de Coimbra, Portugal), foi possível graças à orientação do Doutor Bruno Manadas e do Doutor João Rodrigues.

O trabalho foi financiado por Fundos FEDER através do Programa Operacional Factores de Competitividade – COMPETE e por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., OE FCT/MCTES (PIDDAC) sob os projetos: PTDC/NEU-NMC/0205/2012, strategic project UIDB/04539/2020, POCI-01-0145-FEDER-016428 (ref.: SAICTPAC/0010/2015); POCI-01-0145-FEDER-029311 (ref.: PTDC/BTM-TEC/29311/2017); POCI-01-0145-FEDER-30943 (ref.: PTDC/MEC-PSQ/30943/2017); PTDC/MED-NEU/27946/2017; e pela Rede Nacional de Espectrometria de Massa (RNEM) sob o contrato POCI-01-0145-FEDER-402-022125 (ref.: ROTEIRO/0028/2013).

The present work was performed in the Life Sciences Mass Spectrometry group of Center for Neuroscience and Cell Biology (University of Coimbra, Portugal), and it was possible thanks to Doctor Bruno Manadas and Doctor João Rodrigues.

This work was financed by the European Regional Development Fund (ERDF) through the COMPETE 2020 - Operational Programme for Competitiveness and Internationalisation and Portuguese national funds via FCT – Fundação para a Ciência e a Tecnologia, I.P., OE FCT/MCTES (PIDDAC) under projects: PTDC/NEU-NMC/0205/2012, strategic project UIDB/04539/2020, POCI-01-0145-FEDER-016428 (ref.: SAICTPAC/0010/2015); POCI-01-0145-FEDER-029311 (ref.: PTDC/BTM-TEC/29311/2017); POCI-01-0145-FEDER-30943 (ref.: PTDC/MEC-PSQ/30943/2017); PTDC/MED-NEU/27946/2017; and by The National Mass Spectrometry Network (RNEM) under the contract POCI-01-0145-FEDER-402-022125 (ref.: ROTEIRO/0028/2013).

Dedico este trabalho à minha querida D.Elisabete, a pessoa que me conquistou de todas as bonitas formas possíveis e a quem agradeço o carinho e amor constante, mesmo que longe.

“Reserve your right to think, for even to think wrongly is better than not to think at all.”

Hypatia of Alexandria

o júri

presidente	Prof. Doutor Artur Jorge da Costa Alves Professor Auxiliar c/ Agregação, Universidade de Aveiro
vogal – arguente	Doutor Rui Miguel Pinheiro Vitorino Investigador Principal, Departamento de Ciências Médicas, Universidade da Aveiro
vogal – orientador	Prof. Doutor João Eduardo Aleixo Rodrigues Investigador Doutorado, Universidade de Coimbra

Agradecimentos

Em primeiro lugar, porque este trabalho não teria sido concretizado de outra forma, quero agradecer ao Dr. Bruno Manadas por me ter dado a possibilidade de viver de perto esta experiência na área da investigação. Ao longo desta jornada tive oportunidade de conhecer uma equipa fantástica, divertida, promotora de conhecimento, espírito crítico e que me fez crescer imenso, a nível pessoal e profissional, por tudo aquilo a que me possibilitou. Desta forma, exprimo o meu enorme agradecimento a toda a equipa do laboratório Life Sciences Mass Spectrometry, inclusive à minha colega Margarida, por todos os momentos partilhados.

Em especial, é com enorme estima e admiração por ambos que agradeço aos que me acompanharam de perto, Dr. Bruno Manadas e Dr. João Rodrigues. Obrigada pela orientação, por toda a atenção, ajuda e disponibilidade durante este percurso. Agradeço ainda ao Dr. João as longas horas de partilha de conhecimento, análise e de dúvidas sempre acompanhadas por uma pitada de paciência e amabilidade. À Dra. Cátia Santa, agradeço por se mostrar sempre prestável e atenciosa e por ter permitido também a concretização deste trabalho.

Agradeço ao Professor Luís Souto por ser o meu orientador interno e por se mostrar disponível no início desta aventura a partilhar palavras de apreço da minha prestação como sua aluna para que pudesse alcançar os meus objetivos.

Aos meus colegas de Mestrado de Biologia Molecular e Celular, pelo convívio, pelos trabalhos que realizamos juntos e pelo carinho, obrigada.

Por último, agradeço às pessoas mais importantes da minha vida. A começar pelos meus pais, Maria, António e Jorge, e irmãos, Micael e Luis, que sempre me deram as bases e possibilidades para ir mais longe. Agradeço ainda a eles, à Andreia e em especial às minhas amigas Maria, Sofia, Neuza e Juliana, o apoio e carinho constante, as longas conversas e por me motivarem sempre a alcançar cada etapa. Ao meu colega e agora amigo Daniel, agradeço todos os momentos partilhados a nível profissional e pessoal, todo o apoio e as gargalhadas constantes que alegraram grande parte dos meus dias.

palavras-chave

Doenças psiquiátricas, Esquizofrenia, Proteómica, Espectrometria de Massa, Biomarcadores, Péptidos.

resumo

As doenças psiquiátricas constituem um leque de doenças com enormes incapacidades associadas que podem progredir com o atraso do diagnóstico e com a idade. Muitos são os estudos que procuram dar resposta à fisiopatologia destas doenças, não só com objetivo de compreender melhor estas e solidificar conhecimentos prévios mas também para que a descoberta de biomarcadores se torne alcançável.

A Esquizofrenia é uma das doenças psiquiátricas para as quais o diagnóstico não se rege por deteção de biomarcadores mas sim um conjunto orientações padrão, sintomas e historial clínico, o que possibilita por si só o atraso no diagnóstico. Um diagnóstico correto e precoce é, por definição, sinónimo de melhor qualidade de vida para o paciente. Desta forma, torna-se imperativo a descoberta de biomarcadores que possam diagnosticar e monitorizar a doença.

A espetrometria de massa é uma das técnicas mais aplicadas para pesquisa de biomarcadores. Particularmente, a análise proteómica faz uso dessa técnica para identificar e quantificar proteínas diferencialmente expressas a partir de diferentes amostras biológicas. Desta forma, o presente trabalho permite elucidar de que forma estes resultados podem ser obtidos e, ao mesmo tempo, avaliar se uma análise diferente da convencional, onde o foco reside nos péptidos, permite obter outras informações que auxiliem a estratificação dos indivíduos. Além disto, com objetivo de salientar potencias biomarcadores associados à Esquizofrenia, foi conduzido um trabalho de revisão sistemática que traduz os resultados dos últimos 10 anos de pesquisa neste âmbito.

keywords

Psychiatric disorders, Schizophrenia, Proteomics, Mass Spectrometry, Biomarkers, Peptides.

abstract

Psychiatric disorders are characterized as a range of diseases with associated disabilities that can progress with the delay of diagnosis and with age. Many studies aim to improve the knowledge about the pathophysiology of these disorders, not only with the goal of a better understanding and to refine the previous knowledge but also with the hope that biomarkers discovery become reachable.

Schizophrenia is one of psychiatric disorders for which the diagnosis is not defined through biomarkers detection but use, instead, a set of standard guidelines, symptoms and clinical history that enables the delay of diagnosis. The correct and early diagnosis is, by definition, a synonym of a better quality of life for the patient. Therefore, biomarkers discovery capable to diagnose and monitor the disease are becoming imperative.

Mass Spectrometry is one of the most applied techniques in biomarkers discovery. Particularly, proteomic analysis uses that technique toward the identification and quantification of differentially expressed proteins from different biological samples. Thus, the present work aim to elucidate how these results can be achieved and, at the same time, it will evaluate if a non-conventional analysis, where peptides are the focus, is capable to provide other information that can be useful for patient stratification. Moreover, in an attempt to highlight potential biomarkers linked to Schizophrenia, a systematic review work was performed and the results will elucidate the last 10 years of research in this scope.

Table of Contents

List of figures

List of supplementary figures

List of Tables

List of Supplementary tables

Abbreviations

APs	antipsychotics		**EPS**	extrapyramidal side effects
ACTH	adrenocorticotropin hormone		**EMS**	extensive metabolizers
AC	adenylyl cyclase		**FC**	Fold change
BD	bipolar disorder		**FGAs**	first-generation antipsychotics
BDNF	brain-derived neurotrophic factor		**FDA**	Food and Drug Administration
BMK	biomarkers		**HPA axis**	Hypothalamic-Pituitary-adrenal axis
CNS	central nervous system		**HSP**	heat shock proteins
CID	collision-induced dissociation		**ICD-10**	International Classification of Diseases, version 10
CRH	corticotrophin-releasing hormone		**IDA**	information-dependent acquisition
CSF	cerebrospinal fluid		**LC-MS**	liquid-chromatography-mass spectrometry
DA	dopamine		**LAI-SGA**	long-acting injectable second-generation antipsychotics
D2R	D2 receptor		**MetS**	metabolic syndrome
DRD2	Dopamine D2 receptor		**MVA**	multivariate analysis
DALYs	disability adjusted life years		**ND**	neuropsychiatric disorders
DSM-V	Diagnostic and Statistical Manual of Mental Disorders, edition 5		**NMDA**	N-methyl-D-aspartate
DDA	data-dependent acquisition mode		**3-NT**	neurotrophin 3
DIA	data-independent acquisition		**NIH**	National Institute of Health
DNA	deoxyribonucleic acid		**NNDs**	novel new drugs
EU	European Union		**PMs**	poor metabolizers
ESI	electrospray ionization		**PBMCs**	plasma blood mononuclear cells

PET	positron emission tomography studies		**SCZ**	schizophrenia
PTMs	post-translational modifications		**SPECT**	single photon emission computed tomography
PPIs	protein-protein interaction		**SGAS**	second-generation antipsychotics
PPIs	protein-protein interaction		**5-HTRs**	serotonin receptors
PS	psychostimulants drugs		**SN**	substantia nigra
PQPs	peptide query parameters		**SNS**	sympathetic nervous system
PCA	principal component analysis		**TOF**	time-of-flight
PCs	principal components		**UMs**	ultrarapid metabolizers
PLS-DA	partial least squares-discriminant analysis		**VTA**	ventral tegmental area
RP-LC	reversed-phase liquid chromatography		**VIP**	Variable importance projection
RT	retention time		**WHO**	World Health Organization
SWATH-MS	Sequential Window Acquisition of All Theoretical Mass Spectra		**YLDs**	years lived with disability

CHAPTER 1- INTRODUCTION

1.Introduction

1.1 Neuropsychiatric disorders

Neuropsychiatric disorders (ND) comprise a wide range of mental health disorders that can severely impact the well-being of those affected [1, 2]. These disorders can affect people of different ages and be a main cause of morbidity even in childhood and adolescence [3, 4]. For the current year, there is a mental disorder expected to be the second leading cause of disability worldwide by the World Health Organization (WHO), which is concerning since ND are also a main risk factor for suicide [1, 3]. The effects of these disorders on public health are profoundly negative, and for many reasons, the progress in understanding ND has been slow [1, 5].

In 2010, it was estimated that 15% of the worlds' population, more than a billion people, lived with some type of disability, a 5% increase comparing with previous estimations by the WHO. In the same year, mental and substance use disorders were considered the leading cause of Years Lived with Disability (YLDs), being ND responsible for 31% of YLDs and responsible for 12% of Disability Adjusted Life Years (DALYs) as well. Furthermore, the global burden was established at 10.4%, with mental disorders global direct and indirect cost estimated at US$2.5 trillion, expected to double in 10 years [1, 6-8].

Nowadays, it is estimated that more than 450 million people worldwide live with some form of mental illness, and only in the EU the number of those affected per year is around 165million people [7, 9]. Moreover, one-quarter of the world's population will be manifesting at least one mental disorder at some period of their life [7, 10].

Some of these well represented disabling brain conditions, such as autism spectrum disorder, major depressive disorder, bipolar disorder, and schizophrenia, can affect a significant number of individuals and start early in life [11-13]. Schizophrenia (SCZ) and bipolar disorder (BD) are categorized as chronic and severe ND, and both comprise the top ten of these disabilities [1, 14, 15]. Additionally, individuals suffering from either SCZ or BD have a mortality rate higher than the general population, 2-3 times [14]. Besides these

values, health costs are also substantial. In the year of 2010, a study that included 19 major brain disorders concluded that the total cost of these disorders in Europe was approximately 800 billion EURO, with a cost of 93.3 billion EURO estimated for SCZ [16].

Despite the underpinnings that the field of neurosciences tries to elucidate, the biggest concern with ND is that a continuous lack of knowledge about these disorders makes impossible the establishment of clinical diagnostic tests. The diagnosis, which is based on behavioral markers thorough self-patient report of symptoms, history, and clinical observations, may not be well recognized and, for that reason, not treated properly and at the right time [11, 12, 17]. Thus, the search for biomarkers (BMK) seems to be the method of choice since it can reflect the main changes of the central nervous system (CNS) diseases, namely the dysregulation of molecular expression profiles, and it could hopefully improve the misdiagnosis of patients [3].

1.1.1 Schizophrenia (SCZ)

1.1.1.1 Description and Diagnosis of SCZ

Schizophrenia (SCZ), a complex and severe psychiatric disorder, is characterized by being a highly heterogeneous disorder that affects about 1% of the world's population. The genetic component and environmental factors are usually referred to as two contributors of SCZ, a disorder that can affect people of all ages [3, 18, 19]. According to recent data of a systematic review study from Charlson et *al.* (2018), contrary to other studies, no sex differences were found when compared to gender prevalence [20]. People living with this disorder have an average life expectancy significantly reduced (~) approximately 20 years lower than the general population. Nonetheless, the mortality rates were high across all age groups [10, 20]. The establishment of an SCZ diagnosis can be separated by several years since the first symptoms of the disease and evidences have suggested that early detection can improve the quality of life of patients with psychotic illness [21].

The diagnosis is based on standard guidelines and behavioral reports, clinical history, and observations of patients. The standard criteria are employed by systematic classifications, namely the Diagnostic and Statistical Manual of Mental Disorders, edition 5

(DSM-V), and the International Classification of Diseases, version 10 (ICD-10), published by the American Psychiatric Association and WHO, respectively [3, 22, 23]. SCZ is defined as a spectrum disorder. Hence, there is a wide variation in type, severity, and symptoms that people can experience, which does not make easier the clinicians' evaluations, and based on current diagnosis it can lead to misdiagnosis [22, 24].

1.1.1.2 Symptoms and treatment of SCZ

The symptoms, which arise during adolescence or early adulthood and help to identify the disease at earliest stages, are defined as: (i) positive, such as hallucinations, delusions and thought disorder; (ii) negative, as poverty of speech or alogia, lack of motivation and social withdrawal; (iii) and cognitive as attentional and learning deficits. While positive symptoms can stabilize over the course of the illness, negative symptoms tend to increase and become chronic along with cognitive impairments [25-27].

Psychotic symptoms, which integrate positive symptoms, are a defining feature of SCZ spectrum disorders, and their onset defines the first episode of psychosis [11, 15]. Nevertheless, it also occurs in other psychiatric illnesses, such as mood disorders, and can be observed in other medical conditions [28]. Despite being considered a main feature for the onset and diagnosis of SCZ, psychotic disorders are characterized for an earlier stage, a prepsychotic stage termed as prodrome, which is usually missed by clinicians [29, 30].

The treatment of patients is usually based on antipsychotic medication. After the first successfully employed drug, chlorpromazine, in the treatment of SCZ individuals' positive symptoms in 1952, more drugs were introduced in the following years. The drugs were updated through time and were divided into two categories: first-generation antipsychotics (FGAs), also known as typical, and second-generation antipsychotics (SGAs), formerly known as atypical antipsychotics [31, 32]. Apart from one atypical drug, they act as antagonists of D2 receptor (D_2R), a mechanism that was found to be associated with SCZ in the '60s and turn out to be a general target of all antipsychotics (APs) since the first discovery [31, 33]. However, their activity is also directed to other brain receptors, namely other dopamine receptor subtypes and at serotonin, histamine, acetylcholine and norepinephrine receptors [34].

The dopaminergic dysfunction is a common pathway that seems to lead to psychosis in SCZ patients, and for that reason, the patients treated with the first AP medication only had improvement of positive symptoms. Additionally, these patients had severe side effects when the occupancy of D2 receptors was above approximately 80%, also called extrapyramidal side effects (EPS) [31, 33]. The atypical medication had its design based on clozapine profile, the first SGA that was characterized by lower D2 affinity and more ability to block the serotonin 5-HT2A receptor. Despite the low propensity or absence of EPS and the effectiveness to treat patients that do not respond to other treatments, approximately 30%, clozapine has adverse effects as agranulocytosis, and thus, it is not used as a first-line medication [31, 32]. SGA emerged not only with an associated improvement of EPS, but also an improvement in negative symptoms and cognitive impairments, both responsible for global functioning and outcome disturbs [35, 36]. However, the improvement of negative symptoms by SGA when compared to FGA remains controversial. While some studies have demonstrated that SGAs outperform FGAs in negative symptoms, others failed [31, 32].

A recent follow-up study of Corigliano et. *al*. 2018, showed that the administration of long-acting-injectable antipsychotics in patients with SCZ that previously received a consistent dose of oral medication improved negative symptoms, especially in individuals recently diagnosis (5 years or less of illness), a period believed to be crucial for effective treatment [36]. Moreover, as reported in another study, it was found that long-acting injectable second-generation antipsychotics (LAI-SGA) can be an important strategy to prevent suicide, and more advantages are being described and highlighting this kind of treatment as a powerful strategy to increase patients benefits from treatment [36]. The administration of LAI allows close monitoring of the treatment, with the potential to reduce relapse, rehospitalization, nonadherence, and mortality linked to SCZ [37]. Yet, they remain underutilized and mainly applied in patients with chronic SCZ in order to maintain treatment adherence [38].

For clinical use, the antipsychotic medication should be approved by regulator entities as Food and Drug Administration (FDA), and thus, not all developed AP shown below on the

timeline (**Figure 1**) were accepted. Contrary to what occurred to FGA, SGA did not follow the same trend and end up with more success, with only the antipsychotic amisulpride not being recognized with FDA approval [39, 40].

Recently, following the mechanism of action of aripiprazol as a partial agonist of D_2R, brexpiprazole and cariprazine were also approved by FDA. However, the affinity for D_2R, as well as for serotonin receptors (5-HTRs), is different [41].

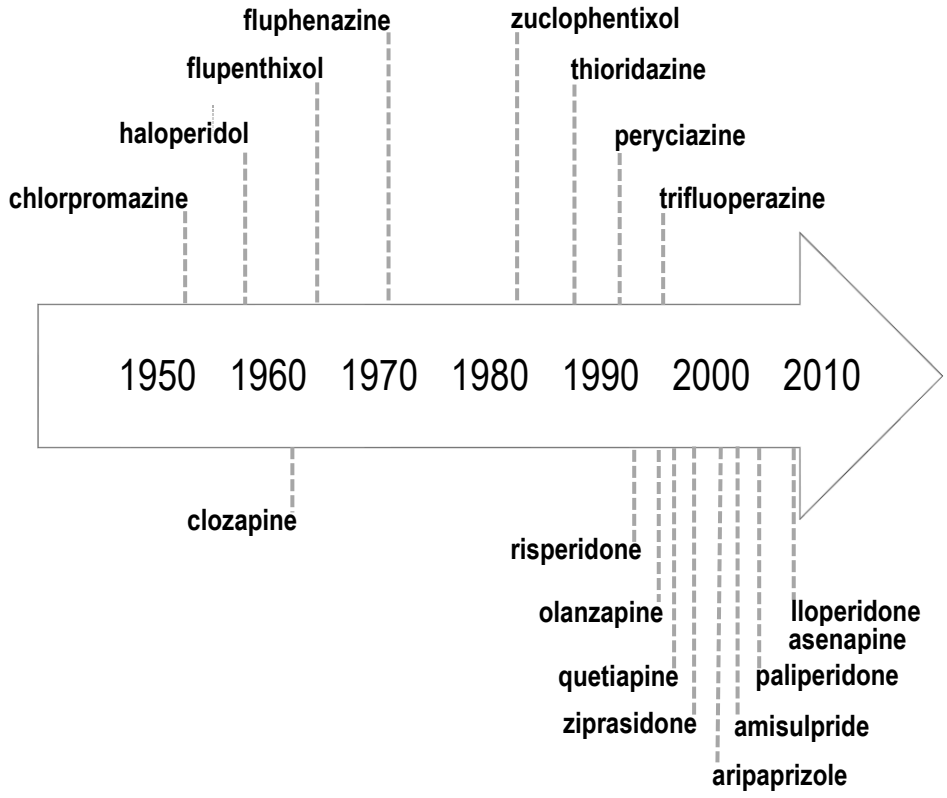


Figure 1. Timeline of antipsychotic drugs. Typical antipsychotic agents are placed on top of the timeline and atypical antipsychotic agents are shown below. Adapted from Shin, J.K., et al. 2011 [42].

The rates of comorbid illnesses associated with SCZ are high, being patients usually linked to an increase of metabolic syndrome (MetS) risk, fixed on 32.5% in SCZ patients in a study of Mitchell *et* al. 2013 [20]. Dysregulated glucose homeostasis and body mass index are common findings in SCZ patients, and despite being found on drug naïve patients, it is

also frequently associated with antipsychotic medication side effects, especially when three or more psychotropic medications are prescribed to a patient [43, 44]. Hence, the cost of treatment can be expensive since psychotropic medication selected for each patient will benefit from add-on pharmacologic prescription for these effects [3].

1.1.1.3 Pathophysiology of SCZ

Genetic risk factors seem to be shared between ND. Notwithstanding, despite being a highly heritable disease, in values between 60-80% as well as highly polygenic, the risk of genetic variants is rare [30]. It is generally accepted that the interaction between risk factors and neurodevelopment triggers the development of the illness, although the time of changes that happen in the brain is controversial. While some evidence of structural imaging, as magnetic resonance imaging (MRI), suggests that schizophrenia arises from changes in early neurodevelopment, other studies propose that dynamic brain changes happen during the onset of psychosis to the transition to active illness. The regions usually studied are the prefrontal cortex and hippocampus, as they are found consistently altered [18, 45]. However, other brain areas have been studied, such as the anterior cingulate cortex, the corpus callosum, and the mediodorsal thalamus [46].

The majority of the studies were performed in brain tissue, mainly at the beginning of proteomics. Therefore, it has been reported that changes are notorious between SCZ and control brain areas [46, 47]. A decrease in brain volume is one of the findings, which is a consequence of the decrement of grey and white matter and can be seen during the time of diagnosis. After diagnosis, while white matter deficits can stabilize or even have improvements in the course of the illness, grey matter tends to get worse as a loss over time progresses. Thus, and supported by twin studies, the white matter may be linked to a genetic risk factor for the emergence of SCZ and not to the effects of the illness itself, which also supports the idea that an abnormal brain development must have arisen many years before the onset [47]. Other structural brain abnormalities reported are the enlarged lateral ventricles and the reduction of the prefrontal lobe and medial temporal volumes. The underlying pathological processes that could explain the progressive changes are still

unknown but may reflect an abnormal brain maturation, anomalies of synaptic plasticity, undesirable stress effects, or other environmental factors [45].

Besides brain changes, distinct processes are described as being changed in SCZ patients, namely in the onset of the psychosis. This interaction with the anomalous neurodevelopment leads to some hypothesis-driven models. One of them is the hypothalamic-pituitary-adrenal (HPA) axis signaling dysregulation. Coupled with the sympathetic nervous system (SNS), the HPA axis mediates the response to stress exposure, which is considered an influencer of many bodily processes. Beyond dysregulation associated with the development of physical illness, the core symptoms of SCZ termed as "pseudostressors" make these patients exposed to a wide range of psychological stressors. Consequently, the secretion of multiple hormones like adrenocorticotropin hormone (ACTH), corticotrophin-releasing hormone (CRH), and cortisol can increase. A couple of evidences usually seen in these patients, as coronary disease, insulin resistance, and lipid abnormalities, have an increment of risk with the prolonged exposure to high levels of cortisol [45, 48, 49]. However, cortisol levels can also have an impact on the immune system, which is also named as a contributor to the development of SCZ [49].

In SCZ patients, it is often observed an increase of pro-inflammatory cytokines with preclinical and clinical literature incidence in IL-6. In the referred context of a lasting stress environment, cortisol can switch its important role as an anti-inflammatory hormone to a responsible for a chronic inflammatory state, developing by turn the immune system glucocorticoid receptor resistant [26, 50, 51]. Nonetheless, the cytokine model of SCZ suggests that prenatal or early life period infection and regular inflammation in adulthood possibly explains the increase of inflammation in the brain of these individuals [51].

The neurotransmitter signaling of dopamine (DA) and glutamate has also been reported as altered in SCZ patients and cytokine dysfunction seems to be linked with it, as well [51]. Comparisons between SCZ patients and neuropsychiatric health controls based on Positron emission tomography studies (PET) show that dopamine contents are different in distinct brain areas. In particular, it is possible to observe a difference in the hippocampus, where the dopamine system presents as overactive in SCZ patients [52]. This

hypothesis model, characterized as the most enduring, proposes that positive symptoms are a result of the increased release of dopamine or hyperactivity and consequently leads to an exacerbated activation of the D2 receptor. Additionally, the negative symptoms and cognitive impairments can be linked, at least in part, with the reduction of D1 receptor activation. The hyperactivity of dopamine is related to the mesolimbic pathway, and the hypoactive transmission of dopamine is associated with the mesocortical pathway (**Figure 2**), although it is also observed in other brain areas [51, 52].

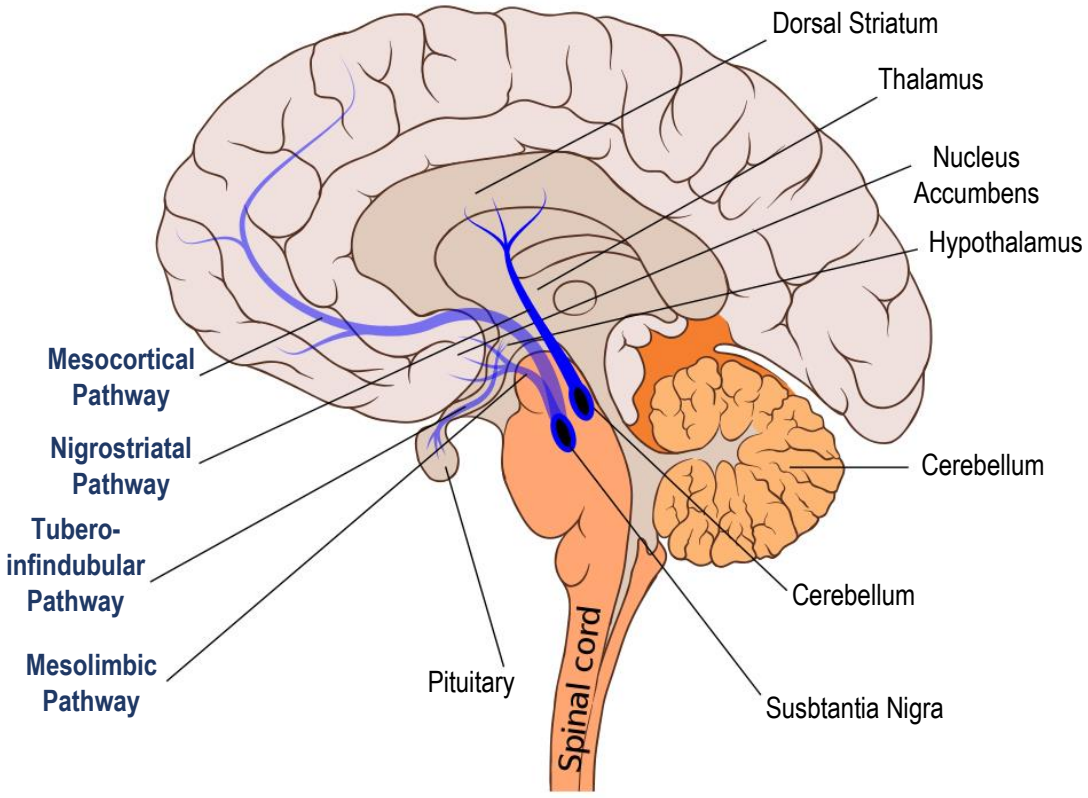


Figure 2. Four main dopaminergic pathways in the brain. The mesocortical pathway projects from the VTA to the cortical regions, especially to the frontal lobes (commonly referred as mesocorticolimbic system). The mesolimbic pathway (reward pathway) connects the VTA to the ventrostriatal areas as nucleus accumbens and is extended as well to amygdala and hippocampus. The nigrostriatal pathway originates from the SN to the dorsal striatum. The tuberoinfindibular pathway is projected from the mediobasal hypothalamus to the pituitary gland. Adapted from Shin, J.K., et al 2011 [42] and Sayin, H. 2019 [53].

In addition to the dopamine approach, the glutamate hypothesis also emerged after being observed that an abnormal function of the N-methyl-D-aspartate (NMDA) receptor, a major glutamate subtype receptor that consequently causes the dysfunction of glutamatergic neurotransmission, induced the typical symptoms of SCZ. NMDAr blockade by antagonists causes the reduction of glutamate transmission, which consequently induces GABA downregulation and DA hyperactivity in DA neurons and GABAergic cells [50, 51, 54]. Together with the dopamine hypothesis, it represents the core mechanisms of SCZ. Both pathways can be altered by pro-inflammatory cytokines, namely by the facilitation of dopaminergic sensitization and by the activation of metabolic enzymes, which have an impact on glutamatergic transmission, e.g., via neuroactive metabolites (kynurenic acid) that act as endogenous antagonists of NMDA receptor [26, 51].

There are other dysfunctions underlying the pathophysiology of SCZ as the energy metabolism (with glycolysis being the central pathway and proteome analysis reports reveals a differential expression of glycolytic enzymes), an imbalance that leads to the existence of oxidative stress events, and a differentially regulated calcium homeostasis and signaling in SCZ patients [18, 46].

1.1.1.4 Dopamine and the dopaminergic hypothesis - Dopaminergic neuronal systems and receptors

Several studies have linked the abnormalities in neurotransmitter systems to SCZ symptoms [55]. The hyperdopaminergic state conduct to the most influential hypothesis about SCZ, which was formulated after the discovery of chlorpromazine antipsychotic action (dopamine receptor blockade) and successive drugs studies. Dopamine is a neurotransmitter that is mainly produced in the substantia nigra (SN) and ventral tegmental area (VTA) of the brain, and it is divided into different pathways (like nigrostriatal, mesolimbic and mesocortical), which are projected from these brain areas [39, 52]. Despite the focus on D2 receptor in SCZ, dopamine receptors or G-protein coupled receptors are divided into two subfamilies, D_1-family (D1 and D5) and D_2-family (D2, D3, and D4) receptors. While D1-family is positively coupled to G protein subunit, D2-family

has an inhibitory effect on adenylyl cyclase (AC) by being coupled to the inhibitory G protein subunit [39].

As the glutamate hypothesis, the dopamine hypothesis was initially based on several indirect sources and evidence that confirmed DA's in similar symptoms to those seen in SCZ, namely in psychotic symptoms. Drug studies had confirmed that some compounds, named psychostimulants drugs (PS) and specifically amphetamines, can increase the levels of dopamine and, consequently, closely resembling the psychotic symptoms that characterize SCZ, while others drugs showed to reduce it by being able to deplete dopamine levels [34, 56]. It was only in the 1970s that the dopamine hypothesis was finally crystallized with observations that the clinical efficacy of AP drugs was directly linked to their affinity for dopamine receptors [34, 57]. In the following years, postmortem studies suggested that neurological changes in the disease were due to an increase of dopamine levels in the striatum and also to an increase in D2 receptor density [34]. Despite the anatomical detail and biochemical specificity that postmortem studies can provide, the confounding effects of AP medication cannot be controlled, which makes it plausible that some of the findings can be considered iatrogenic, as the majority of postmortem brain tissues observed come from patients that have been treated [34, 58]. The increase of D2 receptors levels was then confirmed by studies of postmortem brain tissue and PET of drug naïve SCZ patients and later supported by the fact that all AP drugs have D2-blocking properties [39]. In vivo imaging evidence, such as PET and Single Photon Emission Computed Tomography (SPECT), have been clarifying the dopaminergic function in the brain and refined the dopaminergic hypothesis of SCZ [34]. Moreover, alterations in D2 receptors are supported by genetics findings where notable associations of Dopamine D2 receptor (DRD2) gene as being linked to the etiology and treatment of SCZ [59].

1.1.2 Bipolar disorder (BD)

1.1.2.1 Description and diagnosis of BD

Bipolar disorder (BD), as suggested by the name, is characterized by mood alterations, activity levels, and a broader experience between patients, being categorized

as a "spectrum" disorder. This disorder includes two types: BD-I and BD-II. The first characterized for full and recurrent manic episodes (mood elevation) while the other for the cyclical presence of hypo (a slight and less prolonged form of manic episodes, a milder form of mania)/manic and depressive episodes [60, 61]. The space between these states, depression or mania, is a period of neutral mood state and function, also called euthymia [61].

Typically, BD emerges during adolescence, and the diagnosis relies on the evidence of both acute major depressive and manic episodes, which are distinct. More than a depressive mood, reduced concentration, loss of confidence and energy, diminished interest or pleasure, and frequent thoughts of death are symptoms that characterize the depression state, while extreme peaks of energy, increased self-esteem, and social activity and decreased need for sleep are generally the symptoms observed in manic episodes [61, 62].

This manic-depressive illness with a shift in the mood can affect daily activities performance and the duration of episodes in each mood change between patients, being more emphasized the depressive episodes that had its contribution to the morbidity of the illness [61, 63].

BD has a worldwide prevalence between 1 and 3%, and evidences from different sources suggest that BD is a progressive disease with high morbidity and mortality associated [61, 63]. Even when compared with other serious mental illnesses, the significant risk of metabolic syndrome contributes to these rates [44]. Cognitive impairments are also a central feature in this disorder and can start before its onset, with cognitive functions being more accentuated in advanced stages. Additionally, with the increase of mood episodes, the intervals between them tend to be shorter, which also reflects the less response to the medication [61, 63]. Either way, these clinical outcomes are associated with BD progression [64].

1.1.2.2 Pathophysiology of BD

As for most psychiatric disorders, the etiology of BD is still unknown, but the main possibilities that can contribute to this disorder follow the same line that SCZ, although with slight differences. For example, chronic stressors are present in both phases of BD and, as previously referred, they contribute to the HPA axis dysregulation with a consequent increase of cortisol levels [65]. Also related to the HPA axis, heat shock proteins (HSP) have been studied and associated with its abnormal function since they are involved in cellular homeostasis under stress conditions (stress-responsive proteins). One of the HSP actions is to modulate the conformation of the glucocorticoid receptor and, thus, regulate its function. Alterations in these systems contribute to depression and mood change symptoms, which are extremely related to BD [66, 67].

Changes in the immune system, namely in the release of molecules that contribute to the inflammatory status, are also found among studies with BD individuals. Altered levels of cytokines as IL-6 are reported in these studies, yet in a minor extent than in SCZ individuals [65, 68]. Additionally, a study of Kangguang *et al. (2020)* found that the symptomatic offspring of parents with BD have increased IL-6 levels when compared to the asymptomatic offspring [62].

Studies of neuroimaging show the presence of changes in brain structures like in the prefrontal cortex, hippocampus, and amygdala, although there are no reliable results among them [63, 69]. Correlated with these studies, the most abundant neurotrophin in CNS, brain-derived neurotrophic factor (BNDF), is found with an increased expression on these brain areas that are associated with some of the mood symptoms changing in BD and also associated with its severity. BDNF is altered either in the manic and depression phase and is not the only neurotrophin altered [63, 68]. In both conditions, but with decreased levels, neurotrophin 3 (3-NT) is found elevated in BD patients. This family of neurotrophic factors that had its role in neurogenesis and in the function and survival of neurons also exhibit influence on synaptic plasticity, which is largely associated with neuropsychiatric disorders. Many processes described as altered in BD, as calcium signaling and oxidative

phosphorylation, are important in synaptic plasticity, and usually related proteins are found changed [63, 70].

As reported in data bases, despite the established differences between SCZ and BD, they also share deficits at different levels, such as neurophysiology and brain anatomy, with common and unique genetic features linked to each disease [71].

These serious mental illnesses, composing two of the most commons ND [71], continue to demand in-depth knowledge in order to improve patients' diagnosis and clinical outcomes. Despite the efforts to elucidate the mechanisms or etiology behind neuropsychiatric disorders, they remain elusive and not clarified yet. Thus, the need to search and find reliable biomarkers is becoming imperative.

1.2 The search for biomarkers - omics approaches

To improve the knowledge about these complex disorders, "omics" approaches have risen to shed light on disease pathogenesis and to support a reliable way of prediction and diagnoses for ND [27, 72]. With the fields of "omics", as those who's comprise the central dogma of life, namely genomics, transcriptomics, and proteomics, but not only, the etiology of diseases can be predicted or well understood. The term omics was coined in the nineteen years and has become powerful in the last decades, with omics data about biological molecules revealing specific results and impacting biological progress. With a huge potential associated, high-throughput omics technology can be a solution to predict clinical endpoints, being the ultimate goal of improving patient care and outcome. However, the translation from research to a successful clinical omics-based test is far from the potential of these approaches [73, 74].

The search for candidate biomarkers is the output of "omics" studies. According to the National Institute of Health (NIH), a biological marker, generally just termed as a biomarker, is a "characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [75]. The search for biomarkers in neuropsychiatric disorders started with brain tissue and CSF; nonetheless, the whole body concept emerged and was

established since the integration of the brain and a variety of physiological conditions are reflected in the contents of body fluids. Being one example the "fight-or-flight reflex" [15, 76]. This link created between the brain and the periphery enhance the search for biomarkers in body fluids that could be easily available, like blood (see attached systematic review under submission) [3].

The study about the brain and the associated disorders is complex since it presents a high degree of inter- and intra-cellular heterogeneity and a related proteome due to different cell types and cellular networks. The CNS proteome can change even with minimal alterations in the normal course of its development and/or function [77, 78]. It means that in order to understand the alterations and the mechanism related to a disorder, we should analyze qualitative and quantitative changes in the complete set of proteins that are encoded by an organism genome at different or specific points in time [15, 79].

Proteomics can be a powerful tool since it can give a real-time evaluation of an individual state, health versus disease, and, in an ideal way, predict the susceptibility to develop a specific mental disorder [4, 77]. Hence, in order to improve the knowledge about the range of neuropsychiatric disorders, namely, to elucidate the underlying mechanism of disease, neuroproteomics is the scientific field that aims to study the proteome of CNS [77, 79].

There are some advantages to analyze proteins instead of genome and transcriptome. When compared to the genome, the dynamic and the possibility of not only identify but also quantify the proteins makes the proteomic approach more reliable to evaluate psychiatric diseases at different levels (diagnosis, prognosis, and treatment prediction). Moreover, proteins-based tests can offer the nearest view of the pathophysiological process behind the ND since their expression and function are the result of what happens during post-transcriptional (e.g., alternative mRNA splicing) and pos-translational events (e.g., phosphorylation, glycosylation, oxidation) as well as the interactions between them [3, 4, 80].

1.2.1 The Quest for Biomarkers in Neuroscience

Psychiatric disorders are known for being multifactorial and heterogeneous in expression. Thus, no single molecular event would be enough to explain what happens and what is behind these disorders [26, 81]. Additionally to the high complexity and heterogeneity of those affected, there is a concern in biomarker identification within neuropsychiatric disorders since more than the overlap of symptoms, many candidate biomarkers for a specific ND can present similar patterns with other psychiatric diseases. The lack of knowledge about these disorders' etiology explains why the progress in understanding them has been slow [21, 82].

The discovery of biomarkers in neuroscience is challenging but may help to reveal disease-related alterations and, consequently, improve clinical settings as predict diagnosis, even before the onset, patient stratification, and monitoring of disease progression and treatment [72]. Genomics studies have not answered these questions, and more than be able to find deoxyribose nucleic acid (DNA) variations and differences in gene expression, quantitative and qualitative comparison of proteomes is required to understand a complex disease as SCZ [46]. Proteomics analysis allows to unravel complex proteins networks and signal transduction pathways altered in the disease [46].

Early and correct interventions will improve the outcome of the patients as the switch of medication and either disease diagnosis is common. Therefore, it would increase the quality of life of individuals and reduce the burden associated with psychiatric disorders, as from misdiagnosis, high rates of hospitalization, and treatment expenses, which have a huge impact on health costs [3, 83, 84].

Replace the interview-based methodology and, in a better way, stratify patients into distinct subgroups, would improve the efficacy of the treatment, and also contribute to a decrease of side effects that are highly responsible for the burden [3, 85].

Around half of the patients with psychiatric disorders do not respond to the medication prescribed, and this commonly "trial and error" or "hit-and-trial" procedure/method of treatment make the clinical efficiency weak [3]. Only in SCZ,

treatment is still ineffective to approximately 40% of individuals, and some of them end up quitting the treatment or having severe side effects [86]. The characterization of protein abnormalities by proteomic approach would, therefore, allow the identification of biomarkers that may direct to the use of personalized medicine in mental disorders [48]. Personalized medicine (**Figure 3**) "will address both health and disease and impact on predisposition, screening, diagnosis, prognosis, pharmacogenomics, and surveillance" [87]. Instead of selecting treatment based on phenotype, this emerging approach relies on interindividual biological variability that is predominant among these individuals, who share similar symptoms that can respond differently to the treatment. In order to optimize the therapeutic approach, an effectiveness-safety balance can be mainly achieved through the information of genomics and proteomics and deep knowledge of an individual's own biology [88, 89].

In 2015, 28% or 1 in 4 of the approved novel new drugs (NNDs) by FDA were personalized medicines, and this growing list is only the tip of the iceberg. Strikingly, two of the 13 newly approved personalized medicines include Aristada and Rexulti, used in the treatment of schizophrenia when patients' treatment is influenced by the CYP2D6 biomarker status [90]. This gene, highly polymorphic, can explain why patients are poor metabolizers (PMs), extensive metabolizers (EMs), and ultrarapid metabolizers (UMs), which is explained by its involvement in the metabolism of many important drugs, as numerous psychotropic drugs. Further, CYP2D6 may contribute to SCZ susceptibility, and it has been suggested that it is involved in the biotransformation of neurotransmitters usually linked to SCZ, as dopamine and serotonin [91].

If we look back, in 2005 only 5% of the new therapies accepted were personalized medicines. A big change began in 2014 and continue until the present since more than 20% of these drugs started to incorporate the approved NNDs [90].

According to the WHO, there is one mental disorder expected to be the second leading cause of disability worldwide for the year 2020 [3]. Hence, the need to put into practice this new paradigm is becoming urgent.

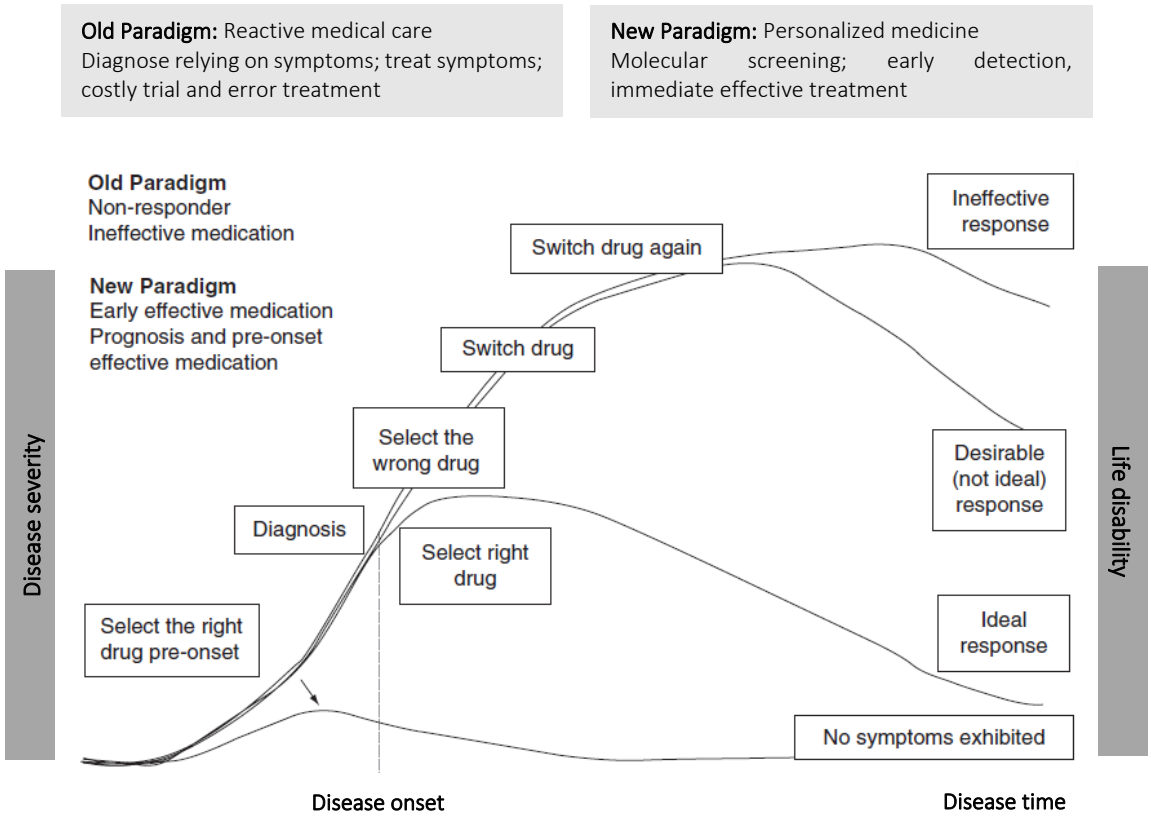


Figure 3. Comparison of the old and new paradigm of medical care as distinguished by the use of biomarkers for improved patient care. In conventional medical practice, a diagnosis based on symptoms and neuropsychological testing will define the treatment of a patient. The selection of the wrong drug and its switch during the course of disease, contribute to the disease severity and life disability. Though, biomarker-based approaches have the potential to change this old paradigm. The use of biomarkers can have an impact on the field of psychiatry by being used in specific and sensitive biochemical tests to follow the traditional questionnaires. The increase of biomarkers tests will allow a massive knowledge of biopatterns in patients that can explain the molecular differences between health and disease state. This, in line with large-scale clinical trials by pharmaceutical companies to improve medical compounds, will help to stratify patients and select the correct treatment. Thus, it will place as quickly as possible the proper patients on the proper treatments trough a more efficient personalized medicine. Adapted from Guest, Guest and Martins-de-Souza. 2016 [3]

1.2.2 Biological markers in neuropsychiatric disorders

Proteomic profiling of ND started with postmortem brain tissue; however, it can only be accessed during autopsies, not being useful for disease diagnosis as not able to be used in longitudinal studies. Additionally, some common variables and confounding factors as postmortem interval and pH range, which can impact the integrity of this tissue, namely,

will contribute to protein degradation, as well as medication and age, consist of the drawbacks that cannot be avoided [15, 92].

Regardless of the disorder been generated in the brain, the composition of blood proteins as other bioactive molecules will reflect those changes since the brain is involved in physiological body functions. This increase concept is known as "fight or flight reflex", making the circulating blood and other biological fluids the sample of choice to analyze protein expression during illness and identify plausible biomarkers [46, 93]. The use of brain tissue is still an option, but the number of studies in body fluids starts to gain some field and to provide evidences that it can be a reliable choice in biomarkers research [27, 46].

Biofluids are suitable samples that can drive to the use of user-friendly tests, being the majority of them of easy access [21, 93]. Based on it, they can be categorized as non-invasive (saliva, sweat, urine, and tears), minimally invasive (blood), and invasive (cerebrospinal fluid).

Due to close proximity to the brain, cerebrospinal fluid (CSF) was considered relevant in the study of brain disorders. This body fluid is rich in molecular entities that can be either the product or the mediator of brain function. Despite the dynamic of CSF, the invasive procedure of sample acquisition by lumbar punction and the minimal amount of material collected limits CSF analyses [15, 72]. As longitudinal studies are essential to elucidate the development trajectories of psychiatric disorders and require different sample collection times, it is important to select more readily accessible and meaningful samples [15].

The view of psychiatric disorders as whole-body diseases made the studies with plasma and serum samples increase over the last years [93]. Besides protein content being significantly richer than what is found in CSF, approximately 500 ml of CSF passes every day to circulating blood [94, 95]. Additionally, dynamic changes can be studied in this type of sample, which can be collected in reasonable amounts and by easy collection [72]. Despite the inherent complexity, a consequence of a high dynamic range, some strategies can be applied to circumventing them [72]. However, some studies also found differences in this highly represented protein content, which can be an advantage when analyzing them. In a

review of Chan et *al*. 2011, which included serum/plasma samples analyzed by liquid-chromatography-mass spectrometry (LC-MS), they were able to differentiate individuals with SCZ from controls and suggested that the majority of potential biomarkers are related to the immune system [84]. Current data indicate that the AP treatment leads to a decrease of cytokines levels; however, if it is a consequence of antipsychotic medication or not, it needs to be elucidated since contradictory findings of an increase and decrease of inflammatory molecules after treatment have been reported [82]. When serum is used instead of plasma, it means that a coagulation process had occurred. This process of proteolytic events can lead to undesirable variability as an incorrect inference of protein content or concentration. Nevertheless, serum samples are not excluded from studies [95, 96].

1.3 Mass Spectrometry

Since its discovery, mass spectrometry-based technologies have been improved and, in the last decades, became a well-suited method for biomarkers discovery, supporting the expansion of the proteomics field [94, 95]. The success of MS in proteomics relies on its specificity and sensitivity (due to advances in LC-MS/MS), giving answers to different proposes [97]. Proteomics tools make possible the qualitative and quantitative, either relative or absolute, analyzes of proteins in different and complex biological samples [96, 98].

These large-scale approaches have the capability to reveal unprecedented insights into the composition, structure, and function of proteins, as well as being able to give information about protein-protein interactions (PPIs) and post-translational modifications (PTMs) modifications [97, 98]. The knowledge about protein expression and modifications, especially in biological processes, can make the bridge to a better understanding of the underlying mechanisms behind physiologic versus pathological cases [77].

1.3.1 Basic principles of MS

Proteomics analysis is a multistep process for which improvement of sample preparation, state-of-the-art mass spectrometry instrumentation, and bioinformatic handling data are required to treat the amount of data that can be revealed [98]. The huge majority of proteomics experiments use two essential types of data, provided for the mass spectrometer, to identify either proteins or peptides.

At first instance, a precursor ion is identified by its mass-to-charge ratio (m/z), which is informative to the molecular mass of the compound being analyzed [95, 99]. This involves three steps: i) molecules are converted to a charged state by the ionization source; ii) ion separation on the basis of their m/z values through the mass analyzer (via magnetic or electric fields); iii) detection of the separated ions as electric charge obtaining signals proportional to the abundance of each species [100, 101].

Second, the use of tandem mass spectrometry (MS/MS) can be feasibly added to generate an MS spectrum. In this MS/MS mode, a precursor ion is selected in the first mass analyzer and subjected to collision-induced dissociation to generate, in the second-stage mass analyzer, a fragment ion pattern. The analysis of this fragmentation pattern can infer the amino acid sequence of the peptide ion but, most importantly, can be used to identify the peptide [95, 100, 102].

1.3.2 The basic components of a mass spectrometer

Of the seven major components that represent the mass spectrometer, shown in **Figure 4**, three of them are considered the main features of this equipment: an ionization source, a mass analyzer, and a detector [77, 102]. Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization are the most common ionization sources. Then, the separation of ions according to their m/z in mass analyzer can happen through the use of quadrupoles or a time-of-flight (TOF) analyzers, among other [77]. When ions arrive at the detector, the kinetic energy is converted into a signal current, and results are displayed as spectra, recorded by the data system [77, 102, 103].

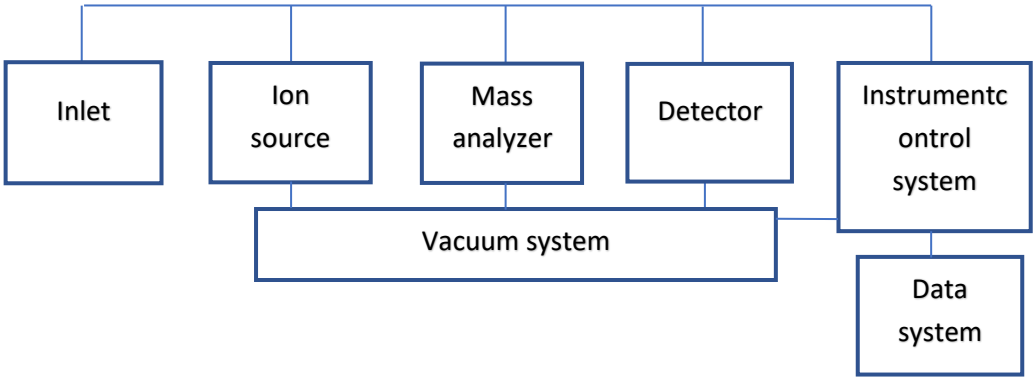


Figure 4. A block diagram of Mass Spectrometer components. The ion source, mass analyzer, and the detector are maintained under vacuum. The instrument control system monitors and controls all parts of the instrument, and the produced data are recorded by the data system. Adapted from Kinter and Sherman. 2005. [102]

1.3.3 "Omics" approaches and mass spectrometry

The measurement of differences between physiological states is one of the most important tasks in proteomics [104]. In the beginning, successes in proteomics approaches were supported by two-dimensional electrophoresis (2-DE), with complex proteins mixtures being separated by its molecular charge (isoelectric point) and mass (molecular weight) in a first and second dimension, respectively, and proteins abundances based on stained protein spots intensities, followed by MS analysis for protein identification [15, 105, 106]. Although improvements were made, other methodologies emerged to sustain some of the previous technicial drawbacks, namely to face the dynamic range limitations and the unsuitable separation and detection of some protein subtypes, as membrane proteins [15, 98].

Nowadays, bottom-up proteomics techniques are the most widespread proteomic workflows [97]. Considered as a robust, fast, large-scale, and high-throughput analysis of the proteome, this approach follows a peptide-protein inference logic, where the sample preparation begins with proteins being extracted and digested by a specific protease, such as (most commonly) trypsin, in opposition to the use of large intact proteins [97, 107]. The resulting peptides are then separated by reversed-phase liquid chromatography (RP-LC),

and before entering the mass spectrometer, peptides are ionized by electrospray ionization [107]. Following this and during the gas phase, peptides ions are fragmented (ex: collision-induced dissociation, CID) to generate MS/MS spectra. Then, mass-spectrometry specific computational pipelines are used for the characterization of eluted peptides [97].

Due to advances in LC-MS/MS techniques, quantitative proteomics can be classified into two major approaches: labeling and label-free techniques [104, 106, 108]. Labeled methods are based on light/heavy peptide intensities and involve different isotope labeling, including chemical, enzymatic, or metabolic, as the most widely used strategies [108]. However, they are expensive and cannot be easily applied to all types of samples [106]. Regarded as a reliable, versatile, and cost effective alternative approach (to labeled quantitation), label-free quantitative proteomics has gained significant interest in recent years [106, 108]. Face to one of its advantages, a high degree of reproducibility is important in this technique since the unlimited number of samples that can be compared are analyzed separately. Hence, it is important to minimize the differences across analysis, which can be achieved through the control of various sources, such as sample preparation process, analytical equipment performance, and low data quality [96]. The quantification is based on MS/MS spectra (spectral counting) or peak intensity measurement (area under the curve) if liquid chromatography is coupled with mass spectrometry [79, 108].

1.3.4 Data Acquisition Methods: DDA and DIA

Distinct acquisition methods are employed for bottom-up approaches, which are dependent on the methodology of the study and the type of instrument available. If the aim is to discover and analyze the proteome within a sample, then it can be selected a data-dependent or data-independent method. On the other hand, if the goal of the study is based on the type of hypothesis-driven, then it will require a targeted approach. The way of precursors are selected for fragmentation and how these ion signals are recorded make the distinction between them [81].

One of the LC-MS/MS strategy's main aims for protein identification, also called shotgun, is based on a data-dependent acquisition mode (DDA, also known as information-dependent acquisition- IDA). Regarding the DDA method (**Figure 5**), the mass spectra of all

precursor ions present in a sample at a specific time are recorded at the MS1 level, followed by the selection of the most intense ones to be fragmented, resulting in the MS/MS (MS2) spectra [79]. Thus, and as the name suggest, there is a previous selection of data before protein identification, with peptide quantification information being reached at MS1 level and peptide identification achieved at MS2 level. Within many instrument configurations, DDA proteomics is dominated by quadrupole-Orbitrap analyzers. However, time-of-flight instruments are also promising [97]. As a consequence of the stochastic nature of the approach, the DDA method also has its limitations. It tends to be biased, with low reproducibility between multiple samples, being significantly affected by sample complexity and dynamic range of analytes [81, 109].

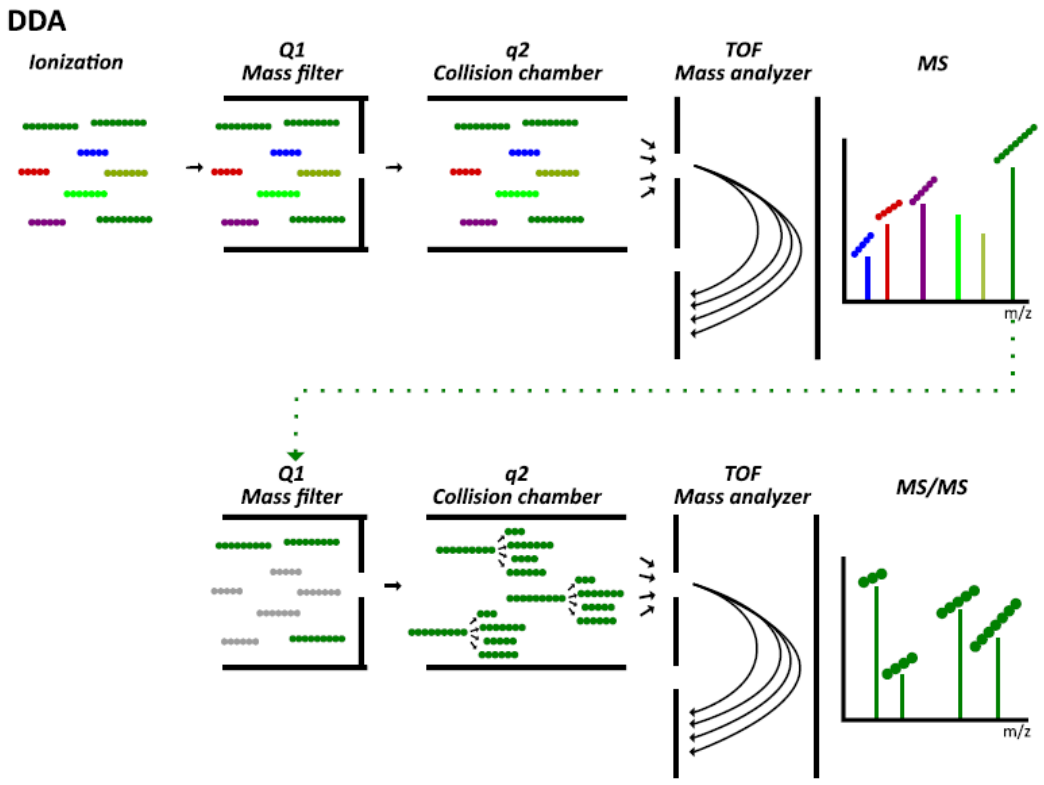


Figure 5. In DDA mode, the first stage of a tandem MS is defined by the recording of all coeluting peptide ions (upper panel). Then, the most intense precursor ions will be fragmented and analyzed in the second stage of tandem MS (bottom panel). Adapted from [109]

To overcome some of the previous limitations, data-independent acquisition (DIA) methods were introduced and have gained some relevance since it can identify peptides

undetected in a typical DDA experiment [106, 109]. As opposed to DDA, DIA methods select the precursor mass range, which is then divided into determined m/z ranges. The information within an isolation window, for all detected precursor ions, will be represented by tandem MS Data. Peptides of interest within this data are then identified by tandem MS spectral libraries [81]. The DIA method can work either as a discovery or pseudo-targeted approach and is generally divided into two distinct groups: (i) those for which fragmentation spectra of the entire mass range are simultaneously acquired (full m/z range), and (ii) those that use sequentially isolated windows to scan the m/z range (selected m/z range) and thus, reducing the complexity of the fragmentation spectra [81, 106, 109]. Despite the fact that it surpasses the DDA method with some methods advantages, DIA also has some issues such as the limited dynamic range of 4 to 5 orders [97].

SCZ biomarker literature shows that several DIA workflows have been used, and they can be distinguished by the way data is collected and analyzed [81]. Sequential Window Acquisition of All Theoretical Mass Spectra (SWATH-MS) is one DIA emerging strategy that has also gained special attention [110].

1.3.5 Sequential Window Acquisition of All Theoretical Mass Spectra (SWATH)

The SWATH-MS approach was developed in 2012 and is a method that allies quantitative consistency and accuracy to a greater proteome coverage capability. Additionally, it is a well-suited method for projects that involve a large number of samples and to evaluate specific PTMs and PPIs, being a good choice for projects that, as biomarker studies, require this type of properties [109, 110].

The particular innovation of this method is based on the proposed data extraction methodology. In a single sample injection, the acquisition of the fragmentation spectra of all precursor ions relies on a defined precursor retention time (RT) and m/z range, which generates complex fragment ion maps. A targeted data extraction strategy is applied to the resulting maps in order to obtain quantitative information of particular peptides, according to the peptide query parameters (PQPs) established previously in spectral libraries. The fragmentation spectra from all precursor ions are accomplished by small and sequential windows of defined size. The confidently identified peptides are the ones that will allow for

protein quantification, which will be determined through fragment ions intensities (MS2 intensity), also named as peak groups, corresponding to each previously identified peptide. Both quadrupole and TOF mass analyzers perform this SWATH-MS data acquisition that typically works with 25Da and a set of 32 overlapping windows covering a total range between 400 m/z and 1200 m/z, with a total cycle time being achieved with approximately ~3.3 s [109-111].

Currently, the SWATH acquisition method can be adjusted for the specific set of samples in analysis through the possibility to select variable precursor isolation windows. It consequently improves the in-set characterization of densely populated mass ranges by reducing the number of selected precursor ions to be fragmented [109].

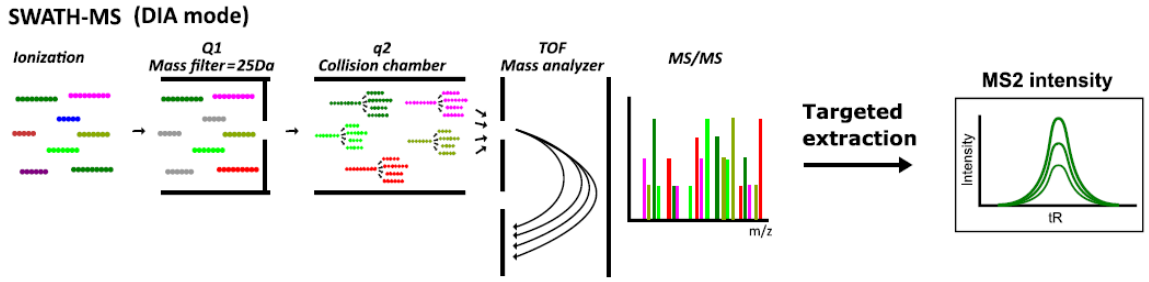


Figure 6. SWATH data acquisition method for MS/MS spectra in a QqTOF system. After ionization, some precursors of a selected mass range are isolated in the first quadrupole (Q1), and then they will enter the collision chamber (q2). Sequential windows of defined size will cover the entire mass range and the confident peptides/proteins are quantified through the fragments ions intensities (MS2 intensity). Adapted from [109]

1.4 Big data

Medical and Science fields are well related to the generation of Big Data, and psychiatry is no exception [112]. Omics biotechnologies lead to the extraction of knowledge from high volume, varied and complex data, intrinsic to Data science, which is known as "big data" [113, 114]. New purposes, clinical and researches applications, have emerged in the proteomics field, in part due to high throughput technologies as MS [112]. Mass spectrometry provides many fine details as insights about protein abundance, expression

patterns, and PTMs that require the need to work with “big data” [115]. For traditional data-processing systems and algorithms, it can be difficult or impossible to process this type of data [116]. Nevertheless, the evolution of computational power, open-source tools, and scripting languages, such as R and Python, are key revolution developments [113]. In addition to univariate analysis, these programs are essential to work with complex datasets, which allow us to apply multivariate analysis (MVA) methods [117]. MVA can be applied for structural simplification or data reduction and grouping datasets by investigating patterns [117, 118]. Two examples of strategies that are used for this are principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) [117].

PCA is a simple MVA technique that allows the reduction of data. This unsupervised projection method uses orthogonal transformation and converts the interrelated variables of a data set into uncorrelated variables. Thus, based upon variances of the first-mentioned variables, a second set of variables is produced, and it can be seen as principal components (PCs) that are ranked in a single analysis [117, 118].

PLS DA, as opposed to PCA, is a supervised method and the goal of this technique is to get a linear regression model through the projection of a predictable variable, matrix Y, from a set of independent variables or predictors (matrix X). This is particularly helpful when matrix X is characterized by a large set of independent variables [117, 119]. The predictive power of this model is obtained through what is called latent variables, extracted from the predictors [119].

CHAPTER 2- AIMS AND OBJECTIVES

2 Aims and objectives

2.1 Aims

Proteomics studies have been growing in psychiatry and show that the use of biological markers can help to improve the knowledge about this complex disease, schizophrenia, by setting the differences in molecular expression profiles between health and disease state. Additionally, it could help during the clinical practice, to establish of a correct diagnosis and patients' stratification. Currently, the use of non-invasive samples (or minimally invasive) and more user-friendly tests support the urgency to put into practice this approach. However, despite the increment of studies, there is still a lack of articles in the literature able to reveal common findings between similar studies (same disease, equal or distinct samples).

Consequently, this work aims to expand what is being recognized in the literature through a systematic review and meta-analysis by using three specific keywords: Schizophrenia, Proteomic*, and Mass Spectrometry. Moreover, since experimental laboratory work was not recommended during the COVID-19 lock-down and restriction measures, a data analysis was performed. From plasma blood mononuclear cells (PBMCs) of patients with schizophrenia and as a complement to previously analyzed data through MS technique, three approaches were studied: i) protein centered analysis; ii) peptide centered analysis and iii) protein vs peptide centered analysis. From this, the goal was to evaluate if specific peptides could improve patient stratification. With peptide centered analysis we confirm not only the trends observed through protein centered analysis but we can also analyze distinct peptides behavior that point to the presence of different proteoforms.

2.2 Objectives

2.2.1 Data analysis- Peptide-centered analysis

The above aim was accomplished by completing the followings tasks:

- Proteomics analysis of PBMCs samples from individuals with schizophrenia (First Episode Psychosis) compared with matched healthy controls

- Summary of peptides that pass the analysis of false discovery rate (FDR)and the relative standard deviation (RSD) threshold between groups – statistical analysis of the variables.
- Evaluation and comparison of Univariate Analysis and Multivariate Analysis
- Summary of proteins based on peptide analysis
- Differences in proteins and peptides behavior
- Discussion of results

2.2.2 Systematic review and meta-analysis (attached in annexes)

The above aim will be accomplished by completing the followings tasks:

- Literature review of mass spectrometry proteomics studies applied to human peripheral fluids in individuals with Schizophrenia
- Use of PUBMED and Web of Science for the computer-based search
- The publication of articles must be comprised between the year 2010-2020
- Data compilation and analysis of the reports
- Application of eligibility criteria
- Selection and analysis of articles eligible for systematic review and meta-analysis
- Summary of biomarker candidates and meta-analysis
- Integration of information and article writing

Chapter 3-Experimental section

3 Experimental section[1]

3.1 PBMCs samples procedures

3.1.1 Sample collection

A set of twelve blood samples was obtained from six first-episode patients (FEP) and six voluntary controls with informed consent and ethical approval of the Faculty of Medicine's ethical committee of the University of Coimbra (CE-122/2015). Peripheral blood mononuclear cells (PBMCs) were isolated from the blood collected into BD Vacutainer® CPT™ mononuclear cell preparation tubes with sodium citrate (BD Biosciences). After collection, the PBMC isolation was performed according to the manufacturer's instruction in the Neurochemistry Laboratory of the Hospital of Coimbra under the supervision of Doctor Inês Baldeiras. After isolation and prior to processing the PBMC samples were kept at -80°C.

3.1.2 Sample preparation

All samples were thawed on ice and sonicated on a SONICS Vibracell 750 W (60% amplitude, 1 second on 1 second off cycles, for a total of 1 minute of sonication), followed by centrifugation at 4 °C for 5 minutes at 5,000×g (Eppendorf®). The supernatants were collected into 1.5 mL LoBind Microcentrifuge tubes (Eppendorf®) and 50 µL of PBS were added to the pellets in each sample tube prior to a second sonication and centrifugation step with the same parameters. The new supernatant was combined with the previous.

3.1.3 Protein Precipitation

Protein precipitation was performed by the addition of methanol (MeOH LC grade, 4× the sample volume), followed by incubation at -80 °C and centrifugation for 20 minutes at 20,000×g at 4°C. Supernatants corresponding to the metabolite fraction were collected evaporated to dryness in an Eppendorf® Concentrator Plus prior to storage at -80°C. A volume of 50 µL of Laemmili Sample Buffer (2× concentrated) was added to the pellets and sonication at 40% amplitude with 1" on 1" off cycles was performed until total pellet dissolution [120].

[1] – All laboratory experiments were previously performed by Catia Santa or otherwise stated.

3.1.4 Protein Quantification

Protein quantification was determined for all samples using the GEHealthcare® 2-D Quant Kit. The assay was performed according to the manufacturer protocol using bovine serum albumin (BSA) as a standard. The absorbance was read at 480 nm on a Microplate Spectrophotometer using BioTek® KCJunior Software. By plotting the standard curve (BSA standards), protein concentration was determined for each PBMC sample (**Supplementary Table 1**).

3.1.5 Gel Electrophoresis

The volume equivalent to 50 µg of protein from each sample was pipetted into new tubes, and a pooled sample for SCZ or control was performed by pooling the volume corresponding to 10 µg of proteins of each biological replicate. The pooled samples are meant to generate the library of identified proteins (IDA acquisition), while the individual samples are meant for the SWATH (relative quantification) analysis.

A solution of 0.1 µg/µL of malE-GFP was prepared by diluting the MBP-GFP in Laemmili Sample Buffer (2× concentrated), and 10 µL (equivalent to 1 µg of MBP-GFP) were added to each sample and pool to be used as an internal standard [121]. Samples were denatured in a thermomixer (Eppendorf®) at 95 °C for 5 minutes followed by the addition of 2 µL of acrylamide per 30 µL of sample for sample alkylation of cysteine residues [122].

3.1.6 Short-GeLC

In order to perform the short-GeLC technique [122], precast Bio-Rad Gels were equilibrated to room temperature, and the required solutions and equipment were prepared. The samples were loaded in a 4-20% Mini-PROTEAN® TGX™ Precast Gels (Bio-Rad) and electrophoretically resolved at 110 V for 22 minutes (instead of the typical 15 minutes) using a Mini-PROTEAN® Tetra Electrophoresis System (Bio-Rad). A volume of 20 µL of Sample Buffer was loaded into empty lanes.

[1] – All laboratory experiments were previously performed by Catia Santa or otherwise stated.

3.1.7 Gel Staining

Proteins were visualized with Colloidal Coomassie Blue G-250 (Thermo Scientific) staining [123]. Then, the gels were rinsed with double deionized water (ddH_2O) and immersed in the staining solution [10% (v/v) of an 85% solution of orthophosphoric acid, 10% (w/v) ammonium sulfate, 20% (v/v) methanol] to which Coomassie Blue (G-250) powder was added. The gels were then placed in a shaker for at least 1h, after which ddH_2O was used to remove background staining.

3.1.8 Gel band processing

After the gels' background was cleared, each lane was sliced using a scalpel blade into three equally sized sections that were then sliced into smaller pieces. Gel pieces were then transferred in a defined order into a 96-MW plate with all the necessary wells filled with 600 µL of ultra-pure LC grade water [124].

The gel pieces were destained with a destaining solution [50 mM ammonium bicarbonate, 30% acetonitrile (ACN)] by shaking in a thermomixer (Comfort, Eppendorf®) at 1050 rpm, 25 °C for 15 minutes followed by incubation in LC grade water for 10 minutes with the same shaking parameters. The process was repeated until no staining was visible in gel bands, followed by dehydration of the gel pieces in the Concentrator Plus (Eppendorf) at 60 °C for 1 hour.

After this destaining procedure, 75 µL of porcine trypsin (Roche) solution (0.01 mg/mL in 10 mM ammonium bicarbonate) were added to each well. The plates were then put for 15 minutes at 4 °C, then 75 µL of 10 mM ammonium bicarbonate buffer was added to each well. The plates stayed at room temperature, in the dark and overnight (≈ 16 hours), so tryptic in-gel digestion could occur. After digestion, the tryptic solution was removed from each well and transferred to the correct 1.5 mL LoBind Eppendorf tube. The peptides were extracted from gel pieces by sequential addition of solutions with an increasing concentration of ACN (30%, 50%, and 98%) in 1% formic acid (FA). In between each solution addition, the plate was placed in the thermomixer at 1200 rpm, 25 °C for 15 minutes, with the solutions collected into the same correct tubes on each step.

[1] – All laboratory experiments were previously performed by Catia Santa or otherwise stated.

3.1.9 C18 peptide clean up

Peptide fractions for SWATH analysis were pooled together, while samples for IDA and library generation were kept in separate tubes. The peptide mixtures were concentrated on the Concentrator Plus at 60 °C until dryness.

To perform C18 solid-phase extraction, the samples were resuspended to a final volume of 100 µL in a solution of 2% ACN and 1% FA and sonicated on a Sonicator with cuphorn (VibraCell 750 watt - Sonics®) at 20% for 2 minutes (1 second on 1 second off cycles). The extracted peptides mixtures were desalted using C18 OMIX tips. Firstly, the tip column was hydrated with 200 µL of 50% ACN and equilibrated with 300 µL of 2% ACN with 1% FA. The peptides were loaded into the column five times, followed by a washing step with 100 µL of 2% ACN and 1% FA. Peptide elution was achieved with 400 µL of 70% ACN and 0.1% FA and the eluates were saved into new 0.5 mL LoBind Microcentrifuge tubes (Eppendorf®). The tubes were then concentrated using the Concentrator Plus (Eppendorf®) at 60 °C until total dryness. All tubes were kept at 4 °C until further use.

The peptide mixtures were resuspended to 30 µL in a solution of 2% ACN and 0.1% FA, and sonicated using the cup-horn at 20% amplitude for 2 minutes (1 second on 1 second off cycles). Centrifugation at 14,000×g for 5 minutes (Minispin, Eppendorf®) was performed, and supernatants were transferred into appropriate vials for posterior LC-MS/MS analysis.

3.2 LC-MS analysis

3.2.1 LC-MS data acquisition

The tryptic peptides were analyzed on a TripleTOF 6600™ mass spectrometer (ABSciex®) in two different acquisition modes: IDA of the three peptide fractions of each pool (Control and SCZ) for library generation; and SWATH-MS, for each individual sample for protein quantification.

The peptide mixtures were separated by LC (nanoLC Ultra 2D, Eksigent® on a ChromXP™ C18CL reverse phase column (300 µm ID × 15 cm length, 3 µm particles, 120 Å

[1] – All laboratory experiments were previously performed by Catia Santa or otherwise stated.

pore size, Eksigent®) at 5 µL/min with a linear gradient of 45 minutes starting at 5% to 30% ACN in 0.1% FA and 5% DMSO followed by a column wash and re-equilibration in a total run of 66 minutes. Peptides were eluted into the mass spectrometer using an ESI source (DuoSpary™ Source, AB Sciex®) with a 25 µm internal diameter (ID) electrode (Eksigent®) [122].

For IDA experiments, the mass spectrometer was set to full scanning spectra (350-1250 m/z) for 250 ms, followed by up to 80 MS/MS scans (100–1500 m/z from a dynamic accumulation time – minimum 40 ms for precursor above the intensity threshold of 1000 counts per second (cps) – in order to maintain a cycle time of 3.5 s). Candidate ions with a charge state between +2 and +5 and counts above a minimum threshold of 10 counts per second were isolated for fragmentation, and one MS/MS spectrum was collected before adding those ions to the exclusion list for 25 seconds (mass spectrometer operated by Analyst® TF 1.7, ABSciex®). The rolling collision was used with a collision energy spread of 5.

For the SWATH-MS based experiments, the mass spectrometer was operated in a looped product-ion mode, based in [111] with the same chromatographic conditions as used as in the IDA acquisition. A set of 168 windows of variable width (containing a m/z of 1 for the window overlap) was constructed covering the precursor mass range of m/z 350-1250. A 50 ms survey scan (350-1250 m/z) was acquired at the beginning of each cycle for instrument calibration, and SWATH-MS/MS spectra were collected from 100–1500 m/z for 20 ms resulting in a cycle time of 3.5 s from the precursors ranging from 350 to 1250 m/z. The collision energy (CE) for each m/z window was determined considering the appropriate collision energy for 2+ ion centered upon this window collision energy spread (CES) was also adapted to each m/z window.

3.3 Data Analysis

A specific library of precursor masses and fragment ions was created by combining all files from the IDA experiments, and used for subsequent SWATH processing. Peptide identification and library generation were performed with ProteinPilot software (v5.1,

[1] – All laboratory experiments were previously performed by Catia Santa or otherwise stated.

ABSciex®), using the following parameters: i) search against a database composed by Homo Sapiens from SwissProt (release in April 2020) and malE-GFP ii) acrylamide alkylated cysteines as fixed modification; iii) trypsin as digestion type; iv) gel-based special focus and v) biological modifications. An independent False Discovery Rate (FDR) analysis, using the target-decoy approach provided by ProteinPilot™, was used to assess the quality of the identifications, and confident identifications were considered when identified proteins reached a 5% local FDR [125, 126].

Data processing was performed using SWATH™ processing plug-in for PeakView™ (v2.0.01, ABSciex®). After retention time adjustment using the malE-GFP peptides, up to 180 peptides, with up to 5 fragments each, were chosen per protein, and quantitation was attempted for all proteins in the library file that were identified from ProteinPilot™ searches. Peptides' confidence threshold was determined based on an FDR analysis using the target-decoy approach, and those that met the 1 % FDR threshold in at least three biological replicates were retained, and the peak areas of the target fragment ions of those peptides were extracted across the experiments using an extracted-ion chromatogram (XIC) window of 5 minutes with 100 ppm XIC width. The proteins' levels were estimated by the summation of peptides area identified for the protein, considering those that pass through the criteria applied [127].

Note: Due to the presence of hemolysis, a control sample was excluded of the analysis since it displayed a profile completely different when compared with the other samples in analysis. In this condition, other blood components proteins will be interfering and masking the PBMCs proteins (see **Supplementary Figure 1**).

[1] – All laboratory experiments were previously performed by Catia Santa or otherwise stated.

CHAPTER 4- Results and discussion: proteomics data analysis

4 Statistical Analysis

By importing the protein and peptides library file, relative protein/peptide quantification using quantitation SWATH-MS data was performed in PeakView. Here it is obtained a spreadsheet with all the data.

Data was exported to spreadsheets where SWATH data is processed (FDR, normalization, RSD% threshold). Statistical analysis was applied to evaluate the differences between 2 conditions (univariate and multivariate analysis). Here, we used a user-friendly, free statistical tool of R program, which will be better explained. A protein and peptide centered analyses were performed to analyze and understand if i) there are proteins that can be used to address differences between the groups; ii) there are peptides that can be used to address differences between the groups, and iii) there are peptides with a distinct behavior of the represented protein and how different of the protein they can be.

4.1 Procedure

The resultant spreadsheet (supplementary information file) that drives this work has information about proteins and all the peptides used in this identification. For each peptide, the values of False Discovery Rate (FDR) can be seen for all of the analyzed samples, i.e., for all six samples that represent the group of Schizophrenia and for each of the five samples that represent the group of control. This spreadsheet comprises information for more than twenty thousand peptides and it was our starting point. From here, the first thing to do was to analyze the FDR. The FDR, defined by Benjamini and Hochberg in 1995 as “the proportion of the rejected null hypotheses which are erroneously rejected” has been used as a criterion to control the proportion of false positives among the interesting or significant results [128]. For example, this FDR filtering reduces the number of false protein identifications caused by modified peptides. The recent study of Boris Bogdanow (2016), shows that it has a huge contribution to the false-positive rate and that false positives related to modified peptides have an impact on protein quantification and are responsible for incorrect protein expression profiles [129]. Hereupon, we applied the FDR cut-off at this point to identify the peptides that have an FDR below 1% in all

samples of each group. Only the peptides represented for an FDR<0.01 in at least three samples of one of the groups, CTL and SCZ, or both were considered for further analysis, which is a few more than six thousand peptides.

A new spreadsheet was created with this data, and the area of each peptide was calculated. With this aim, the data of ions area were summed for each peptide (see table 1.), for the different samples in the analysis. The next step allowed the normalization of data, being used for this the total ions' area. With normalization, we will be considering the bias and make samples more comparable [130]. To reach this, the area of all peptides was summed for each sample and then, the area previously calculated for each peptide (sum of ions area) was divided by this total of summed areas. After that, we were able to calculate the median, average and standard deviation in the excel spreadsheet for each peptide. With these values achieved, we could find the relative standard deviation (RSD), once again for each peptide.

Table 1. Data of ions area for the different peptides and proteins in the analysis. This table shows two peptides of the same protein. Each peptide has the value of m/z for the precursor ion, the number of precursor charge, the retention time (RT) and a column with one of the samples in the analysis (CTL45) where it is expressed the values used for the sum of areas of each peptide.

Protein	Peptide	Precursor MZ	Precursor charge	RT	CTL45
sp|P35579|MYH9_HUMAN	LQQELDDLLVDLDHQR	650.67	3	3.78	23300778
sp|P35579|MYH9_HUMAN	LQQELDDLLVDLDHQR	650.67	3	3.78	2104528
sp|P35579|MYH9_HUMAN	LQQELDDLLVDLDHQR	650.67	3	3.78	766149
sp|P35579|MYH9_HUMAN	LQQELDDLLVDLDHQR	650.67	3	3.78	176946
sp|P35579|MYH9_HUMAN	LQQELDDLLVDLDHQR	650.67	3	3.78	1047546
sp|P35579|MYH9_HUMAN	VISGVLQLGNIVFK	743.95	2	49.91	4406107
sp|P35579|MYH9_HUMAN	VISGVLQLGNIVFK	743.95	2	49.91	2533416
sp|P35579|MYH9_HUMAN	VISGVLQLGNIVFK	743.95	2	49.91	2291115
sp|P35579|MYH9_HUMAN	VISGVLQLGNIVFK	743.95	2	49.91	-
sp|P35579|MYH9_HUMAN	VISGVLQLGNIVFK	743.95	2	49.91	549619

The RSD is represented by the formula: $s / |\bar{x}| * 100$, where "s" is the value of sample standard deviation and $|\bar{x}|$ the value in module of the sample mean, which is expressed as a percentage and thus, it is multiplied by 100. This variable, a type of coefficient of variation, let us know if the data are tightly grouped around the mean. When compared to the mean of the data in the analysis, the RSD will show us if the data are more spread out or close to each other. After the calculation of RSD, a threshold of RSD < 30% was applied for all the peptides in the analysis. We could notice that most of the peptides that did not pass through this criterion were the peptides that had not significance for one of the groups. Next, a spreadsheet with the peptides that pass through the previous criteria applied as FDR, normalization, and RSD was created. In this new spreadsheet, we calculated or determined the fold change (FC) value for each peptide (more than three hundred and fifty). The fold change will describe the ratio between the cohort group (SCZ) and the healthy ones (CTL). Thus, we will be able to see the difference in protein expression

between these two groups. To do so, we used the mean values of each group (which express the areas detected for each sample) to achieve the FC for each peptide. Below (**table 2**), an example is shown for a specific protein that reached this point of analysis, named Apolipoprotein B (APOB). There are only three peptides in the previous spreadsheet and thus, these three were considered for the FC of the protein. The peptides' area values are expressed at the column of each sample (SCZ group), and they are summed to achieve the area value of the protein for the different samples in the analysis (SCZ4, SCZ5, SCZ6, SCZ8, SCZ9, SCZ12).

Table 2. The protein, Apolipoprotein B, is represented at this point of analysis by three peptides. Only the area values for the group of Schizophrenia is here represented. To reach the value of FC, the area values of each sample (SCZ4, SCZ5, SCZ6, SCZ8, SCZ9, SCZ12) are summed and compared with the same calculation obtained for the group of control.

Protein	Peptide	SCZ	SCZ5	SCZ6	SCZ8	SCZ9	SCZ12
sp|P04114|APOB_HUMAN	IAIANIIDEIIEK	2.17E-05	1.20E-05	1.46E-05	2.01E-05	1.20E-05	1.77E-05
sp|P04114|APOB_HUMAN	LSNDMMGSYAEMK	4.62E-06	2.28E-06	2.10E-06	3.39E-06	3.83E-06	2.97E-06
sp|P04114|APOB_HUMAN	LSQLQTYMIQFDQYIK	1.23E-05	6.90E-06	1.33E-05	1.32E-05	1.50E-05	1.16E-05

After doing the same for the control group, it was possible to achieve the FC of each peptide by doing the ratio between the median of SCZ and the median of the control group. Then, we aimed to do the same to the proteins for which the peptides represent. A new spreadsheet was created and the FC was calculated. Once again, to achieve the FC of each protein at this point, we had to calculate the median of the area values for the group of Schizophrenia and for the control group, calculating later the ratio between them **(see table 3**).

Table 3. The protein here represented is Apolipoprotein B and the respective peptides. For each, the values of median, average, standard deviation (SD) and fold change are shown (FC). For the group of Schizophrenia (SCZ) and control (CTL), the values shown are the result of all samples that make up each group.

	Median (SCZ)	Average (SCZ)	SD (SCZ)	Median (CTL)	Average (CTL)	SD (CTL)	FC (SCZ/CTL)
Protein sp|P04114|APOB_HUMAN	3.16E-05	3.16E-05	6.11E-06	3.67E-05	4.91E-05	3.64E-05	**0.861**
Peptide							
IAIANIIDEIIEK	1.62E-05	1.64E-05	4.15E-06	1.70E-05	2.45E-05	2.12E-05	**0.951**
LSNDMMGSYAEMK	3.18E-06	3.20E-06	9.56E-07	3.97E-06	6.32E-06	6.22E-06	0.801
LSQLQTYMIQFDQYIK	1.27E-05	1.20E-05	2.77E-06	1.46E-05	1.82E-05	1.01E-05	0.872

In a new spreadsheet, one more criterion was applied to search for the peptides with modifications. Modifications of peptides are a common finding in proteomic samples and are responsible for a huge fraction of assigned spectra and/or can be erroneously assigned to incorrect amino acid sequences, which can result in false protein identifications [129]; thus, they are generally not considered in traditional proteomics workflows. Some modifications occur in vivo, while others are introduced in vitro when samples are prepared for proteomics analysis [129]. Acetylation, methionine oxidation, glycosylation, ubiquitination, formylation and phosphorylation are commonly seen and are naturally introduced [131]. Acetylation and methylation, for example, can be a result of both [129]. The traditional determination of protein fold change is based on peptides summation that excludes the modified peptides for each analyzed protein. At this point, we searched for the peptides with modifications, and in a new spreadsheet, they were excluded. After this, we could calculate the new FC for each protein that now did not include the areas of the modified peptides and see the ratio between the FC of proteins that include all the peptides vs the fold change of proteins without modifications.

To keep going with this work and perform the statistical analysis, the R software was used to support the following analysis, specifically the R studio, which integrates the R software environment. The data contained within the table previously analyzed is extracted to a format that R studio is capable of importing and read. First, normality tests

were applied to evaluate the distribution of data. As we had in analysis a small number of samples, it was not expected that data was well-modeled by a normal distribution. A Quantile-Quantile plot (Q-Q plot) was created for a visual assessment of normality (see **Figure 7**). It can work as a prior visual inspection for this inference. However, this data did not provide clear and useful information about sample distribution.

Due to the small number of samples analyzed, the statistical tests applied were non-parametric.

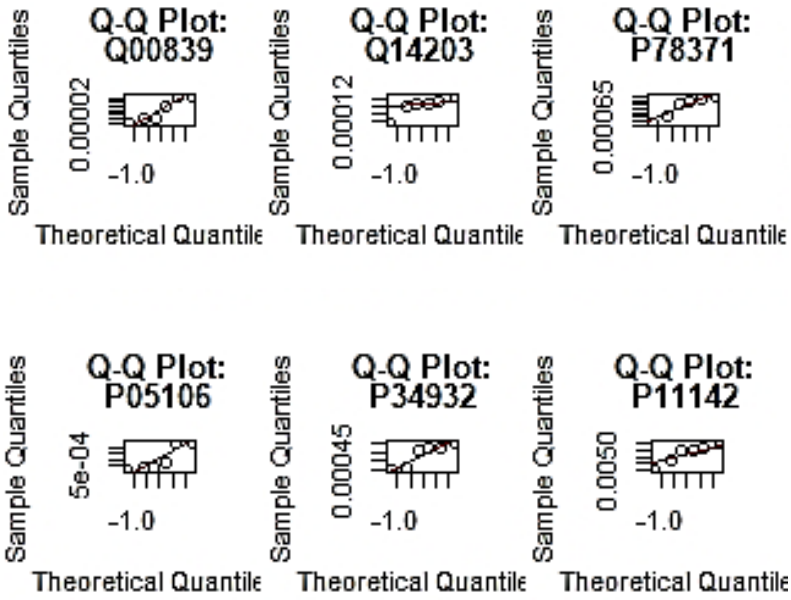


Figure 7. Quantile-quantile plot. On the x-axis of the graphic is plotted the theoretical quantiles, also known as the standard normal, while on the y-axis are the ordered values of our samples (SCZ cohort). The points, which represent the samples in analysis, are not aligned on the standard normal variate, above the straight line. However, this analysis is not enough to provide information about the distribution of data. Generated using R-Studio.

Numerical measures of shape, as skewness and kurtosis, can be used to analyze and provide convincing evidence of data's normality. In this analysis, both coefficients are also suggestive of a nonnormal distribution. Moreover, more robust normality tests, like Shapiro-Wilk and Kolmogorov-Smirnov tests, were also applied.

The Wilcoxon signed-rank test, also known as a distribution-free test, was used as a non-parametric approach [132]. It was used for the comparison of both groups, disease

and control, as this test provides information about the median of samples in analysis and the respective p-values. The median values were previously calculated and used to obtain the FC of each protein. Thus, from the test results, we take each protein's p-value information to the excel spreadsheet.

After this, it was given attention to multivariate analysis (MVA). MVA has multiple advantages. One outcome can be achieved from multiple dependent analysis, which allows us to make a comprehensive analysis and realize how variables relate to each other. For complex datasets, particularly, univariate analysis alone can be insufficient and unsatisfactory. Thus, we can use multivariate data analyses to improve statistical analysis and, in some cases, to contradict the results of univariate analysis [133]. Hereupon, we also aimed to look at the results from both analyses and formulate our own opinion.

There is something that is highlighted as a result of multivariate analysis and that we are not able to recognize with the univariate analysis. Using the first mentioned, we can do a separation between the case/disease and the control group while, at the same time, we can analyze and recognize which proteins promote that separation and classify the dataset. Thus, distinct MVA methods are required for visualization and a type of interpretation that UVA, alone, is not capable of. With that aim, one of the methods employed was the partial least squares-discriminant analysis (PLS-DA). In this approach, as the name suggests, a PLS regression model is applied to our variables [134], which are the proteins (n= 1106, proteins without modifications) or peptides (n= 3534). Using PLS, our aim was to find components from the variables that best predict our groups. In **Figure 8**, a good and clear tendency of separation is observed between the two groups of samples in analysis, control and disease. To analyze this and work with the best predictive power, the model use orthogonal factors named latent variables extracted from the variables or predictors [119]. Some parameters as R2X, R2Y and Q2 are analyzed and give information about the quality of the model [135]. X represents our variables (proteins), while Y is a vector representative of the groups. Thus, looking at R2X and R2Y values, it is possible, by assumption, to see how well the model fits the X and Y data [135]. Q2 represents the accuracy of the model prediction. When higher, a better predictive ability is achieved for the model [135]. This way, we can control and avoid overfitting by choosing the number of

latent variables or components we want to consider. With this in mind, two latent variables were used in this PLS-SA model. In summary, in **Figure 8**, it is shown a PLS-DA model with two latent variables enabling a clear separation between the two groups in study, control and disease. The PLS-DA model is characterized by R2X and R2Y values of 0.557 and 0.887 which represents the fraction of the variation in the X and Y matrix, respectively, explained by the model. Q2Y, with a value of 0.401, is the fraction of variance of the Y variables predicted by the model. RMSEE indicate the Root Mean Squared Error of Estimation.

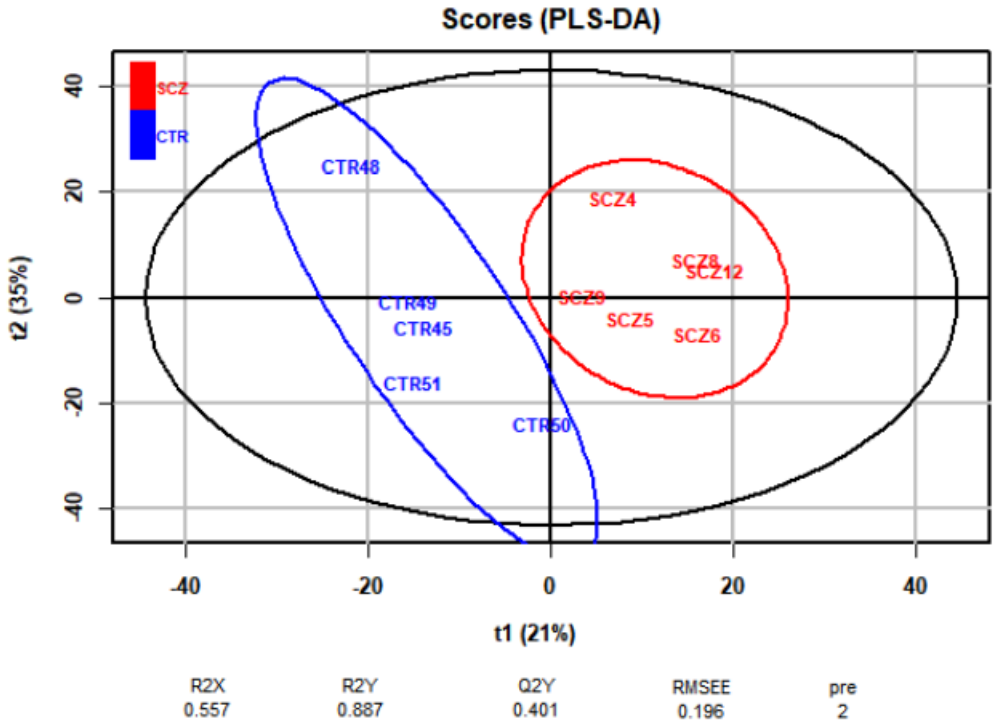


Figure 8. PLS-DA scores plot based on proteomic data. The figure shows the analysis of all samples from both groups, established with the total of proteins identified. It is clear a good separation between control and disease group. 21% of the variance is explained by the component 1 (t1), while component 2 justify in a better way the clear separation between the groups, with 35% of variance explained (t2). R2X: the explained variation of X matrix, goodness of fit; R2Y: the explained variation of Y matrix; Q2Y: goodness of prediction; RMSEE:Root Mean Squared Error of Estimation; pre: predictive component. Data generated using R-studio.

The importance of each variable (proteins or peptides) to the distribution of the scores (groups separation) is expressed in the loadings weights matrix. In **Figure 9**, a 3D-plot was created showing the loadings weights plot of the two latent variables (x- and y-

axis) and the variable importance in projection (VIP) of each variable (z-axis). A color scale is used based on the VIP values for each variable, with the more relevant variables being shown in red.

With the VIP values, it is possible to notice each variable's weight (protein) in the projection of the PLS-DA model [136]. The VIP values are usually used for variable selection. The typical rule is to select the variables with scores greater than 1 [137]. Thus, when we take this into account, we are looking for the relevant proteins.

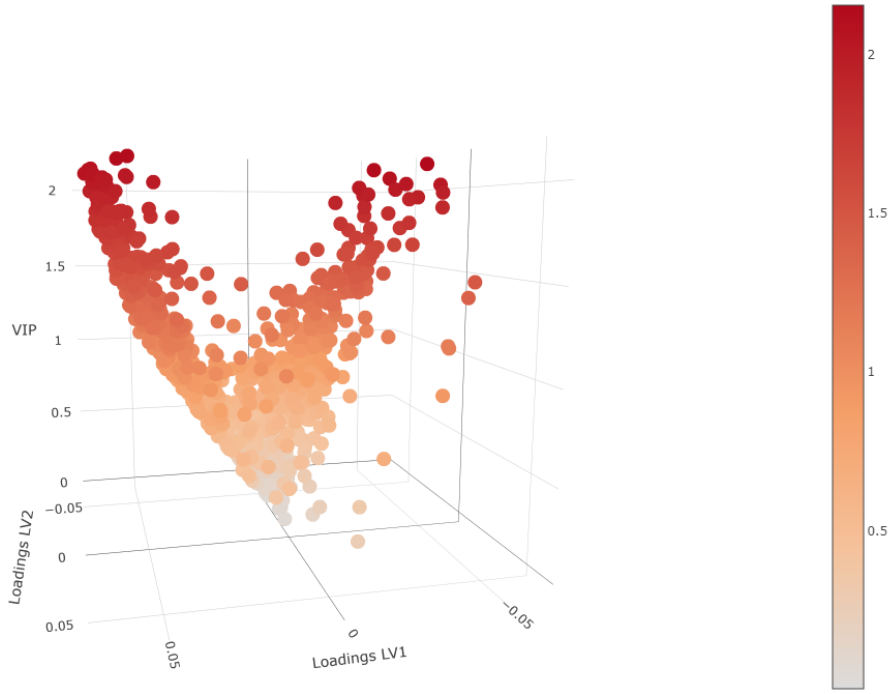


Figure 9. PLS-DA loadings weights plot based on proteomic data. Each point of the plot represents the variable importance in projection (VIP) values for each protein. Relevant proteins are highlighted with dark red color. From VIP value 1, all proteins are considered relevant for the separation of disease and control group. Generated using R-studio.

The results of VIP values were imported into the excel spreadsheet. It was then possible to analyze if the proteins identified by the p-value in the univariate analysis were also considered relevant in the multivariate analysis. To see this, a Venn diagram was

designed, and we could see that all the proteins identified in UVA were also identified in MVA. There was no loss of information (see **Figure 10**).

Of a total of 1106 proteins, only 396 (**Figure 10**) were considered relevant for the model that aim to do the separation of the groups in analysis. The two analysis complement each other since VIP analysis make a selection of proteins that, even without statistical meaning, can be relevant for the groups separation. Thus, despite being less restrictive, VIP analysis allow the identification of variation trends.

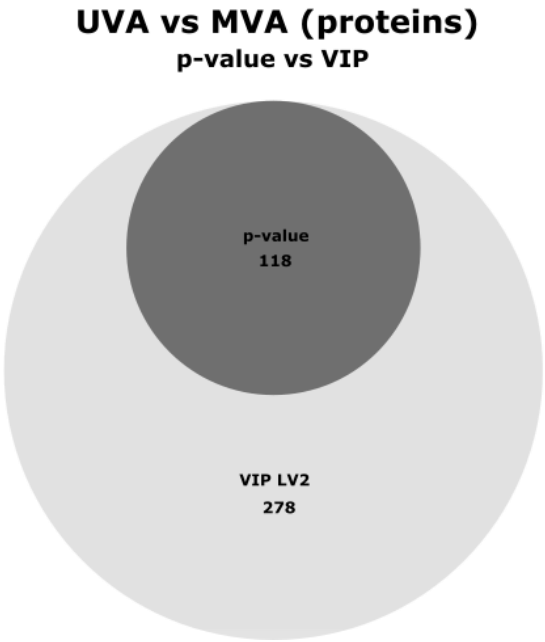


Figure 10. Proteins Univariate Analysis vs Multivariate Analysis. In this Venn Diagram with a total of 396 proteins, we can notice that the 118 proteins that pass through the p-value analysis in UVA, were also selected through VIP analysis from MVA in R-studio. Additionally, there are 278 proteins that were considered as important for the separation of groups only through VIP analysis.

To perform a biological analysis of these 118 proteins and analyze what kind of pathways could be enriched, it was used the Reactome Software. Many proteins were found enriched through this tool and different pathways were highlighted (**Figure 11**). The Supplementary Table 3 shows the statistically significant enriched pathways, with the total

of entities found and the respective p value and FDR value being shared.

Of the pathways considered as altered, they are linked to hemostasis functions with involvement in platelet (see also **Supplementary Figure 2**) and coagulation cascade functions. Thus, to have more than an overview, The Kyoto Encyclopedia of Genes and Genomes, KEGG, was used to promote a better perception of the proteins involved in the platelet activation and complement and coagulation cascade **(Figure 12)**.

As it was previously reported, alterations in immune system, namely an increase of pro-inflammatory cytokines, is often observed [51]. Immune abnormalities have been recognized in SCZ patients in the periphery and CNS and it seems independent of AP medication [138, 139]. Moreover, a meta-analysis of studies performed in blood and CSF, with many studies analyzing PBMCs, identified some cytokines that could be used as state and trait markers [139].

Additionally, dysfunction in the coagulation and complement system in individuals with SCZ have been studied in the last years. A recent study involving individuals with SCZ reports that tissue plasminogen activator (tPA) activity can be affected by different conditions, with elevated cytokine levels being characterized as one. These patients presented low levels of tPA and chronic warfarin therapy was linked to long-term remission of psychotic symptoms [140]. Another recent published study found alterations in complement and coagulation cascade in 11 years old children, which preceded the development of psychotic disorder [141].

This functional succinct analysis shows the ability to identify biological alterations that are specific for the disease. Moreover, with this approach we can reach altered proteins that are biological relevant.

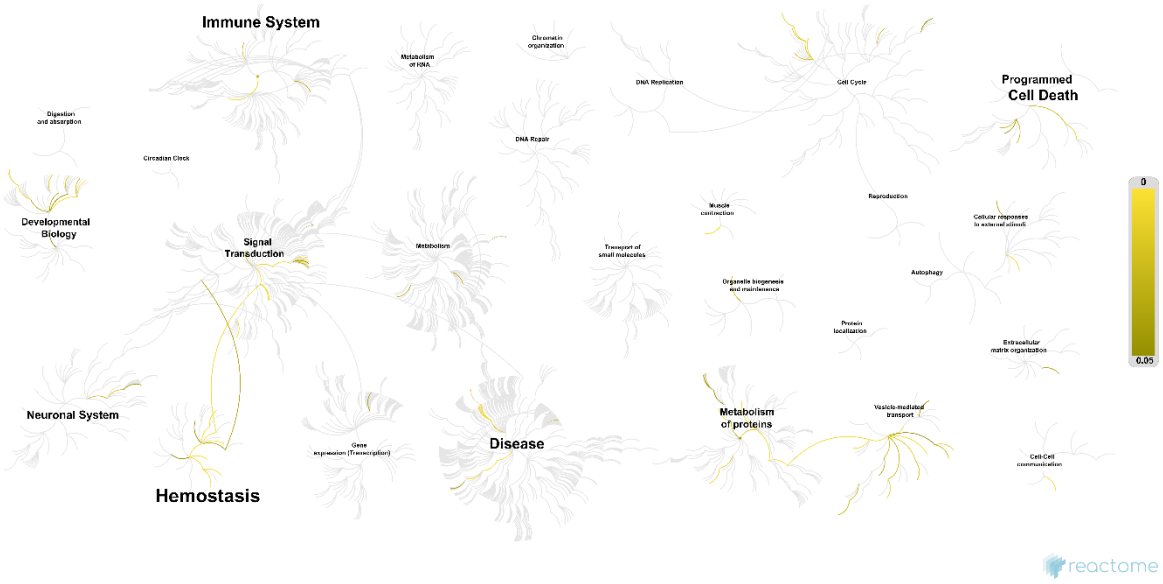
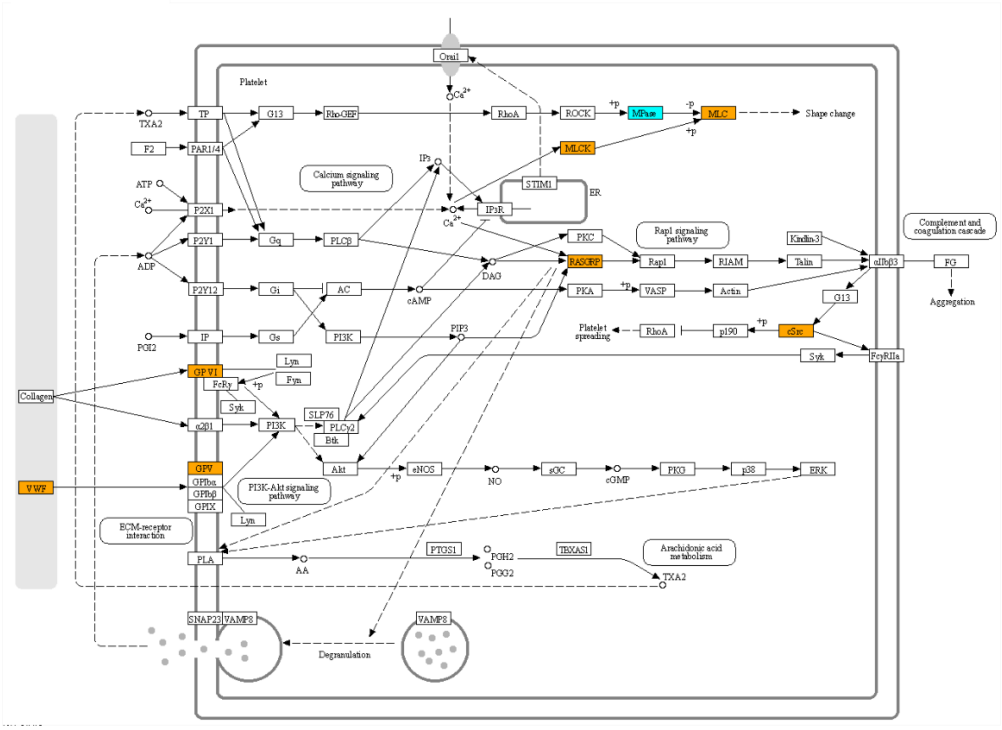


Figure 11. Reactome Genome-wide overview of the 118 proteins reached through proteins analysis and with p-value <0.05 for CTL vs SCZ comparison. The color code indicates the pathways representation based on p-value. Only branches considered significantly enriched by these 118 proteins are depicted in yellow. The pathways that are not significantly over-represented appear with grey light. Image generated from Reactome Software.

A) Platelet Activation



B) Complement and coagulation cascades

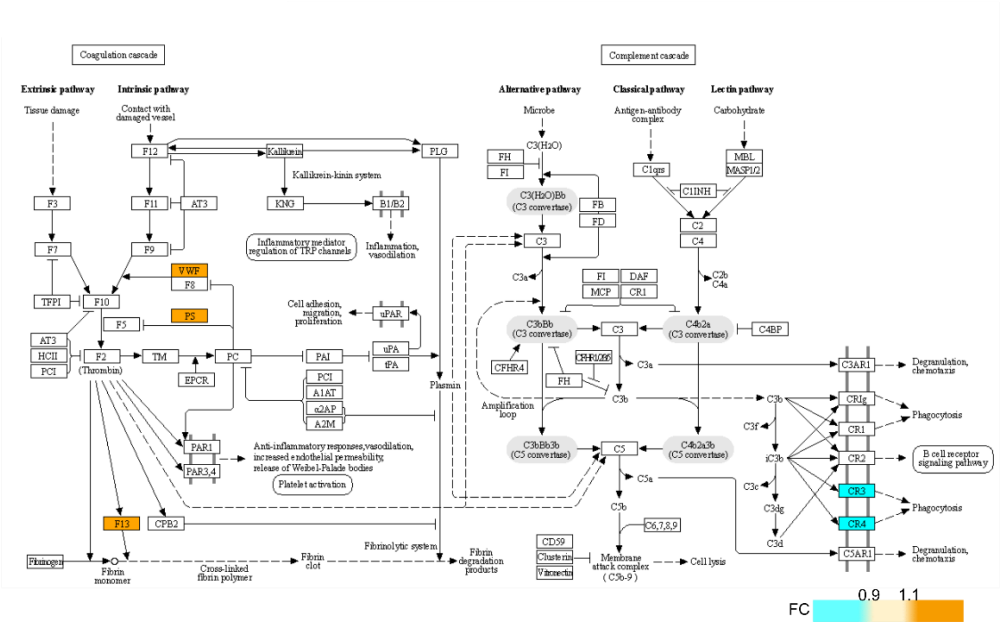


Figure 12. Schematic view of the altered proteins represented in platelet activation (A) and complement and coagulation cascades (B). KEGG Search & Color pathways graphical visualization of platelet activation (A) and complement and coagulation cascades with the indication of the altered proteins where a color code was used to represent up- or down-regulation of the proteins.

For the peptides that reach this point of analysis, a PLS-DA model was also performed. Like protein results, when performing peptide-centered analysis of the data,

we can see through PLS-DA model a clear separation between the groups in the analysis (**Figure 13A**). For this model, the PLS-DA is characterized by R2X and R2Y values of 0.486 and 0.926, respectively, while Q2Y (which characterize, when higher, a better model prediction) has a value of 0.514.

The percentage of variance explained by each latent variable has the same weight. Thus, both contribute equally to the separation of the groups. Additionally, through VIP analysis, we can see, once again, which variables have a huge contribution in this separation, as for the projection of the PLS-DA model (**Figure 13B**). Therefore, of a total of 3534 peptides (1119 proteins), 1338 achieved VIP scores greater than 1, and these are considered the relevant peptides to the establishment of group separation.

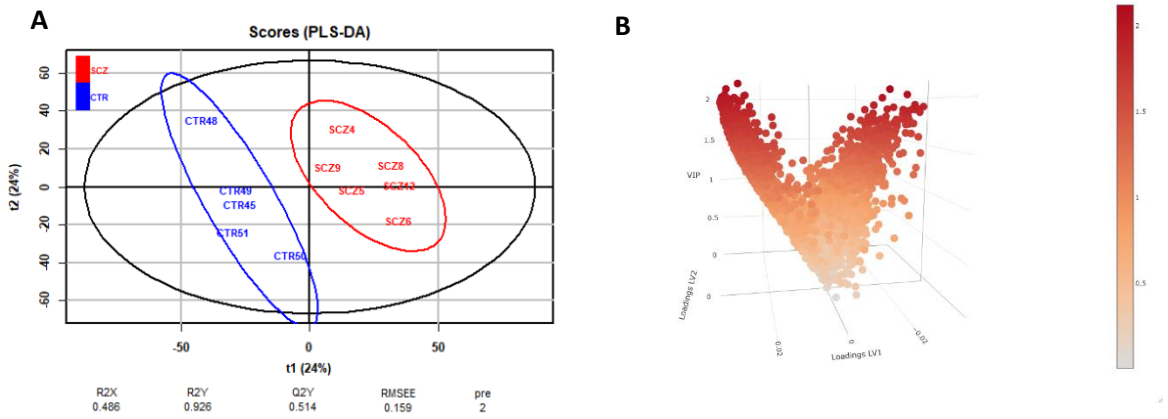


Figure 13 PLS-DA model based on peptide centered data of all samples from both groups, with the total of peptides identified. **A)** PLS-DA scores plot, showing a clear separation between control and disease group. Each component, component 1 (t1) and component 2 (t2), explain 24% of the variance. **B)** Loadings weights plot of the two latent variables calculated for the PLS-DA model, colored by the values of variable importance to projection (VIP). Relevant proteins peptides are highlighted with dark red color. From VIP value 1, all proteins and peptides are considered relevant for the separation of disease and control group. R2X: the explained variation of X matrix, goodness of fit; R2Y: the explained variation of Y matrix; Q2Y: goodness of prediction; RMSEE: Root Mean Squared Error of Estimation; pre: predictive component. Data generated using R-studio. Data generated using R-studio.

Within the total of peptides considered by PLS-DA (representing a total of 578 proteins), 533 peptides (corresponding to 258 proteins) were found through p-value and

VIP analysis (**Figure 14**). Working as a complementary analysis, 805 peptides were also considered important through VIP analysis.

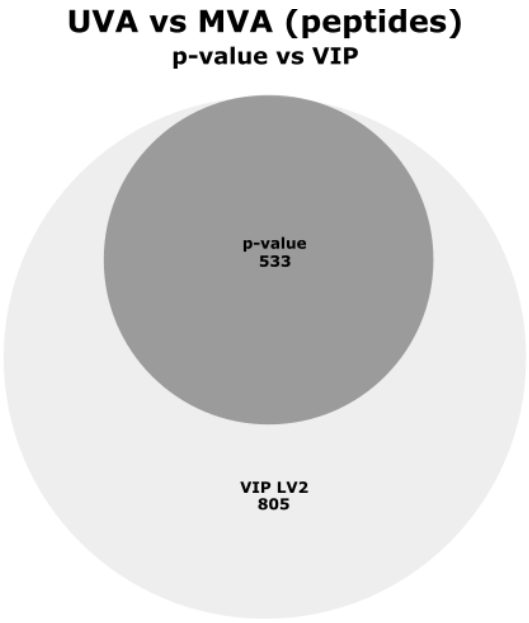


Figure 14. Peptides Univariate Analysis vs Multivariate Analysis. In this Venn Diagram we can notice that the 533 peptides that pass through the p-value analysis in UVA, were also achieved through VIP analysis from MVA in R-studio. Additionally, there are 805 peptides that were also considered as important for the separation of groups through VIP analysis.

Besides being essential to reach this work's aim, the procedures shown before illustrate how VIP analysis can be useful when we want to characterize the groups in analyzes through variables. Some of the peptides considered relevant for this separation and not considered through UVA have a value of p-value close to the statistical meaning.

As the biggest goal of this analysis was to focus on peptide centered analysis, this approach allowed the selection of important variables analyzed until this point.

CHAPTER 5- Results and Discussion: indirectly searching for proteoforms

5 Results and discussion

The present analysis used 11 PBMCs samples comprising 2 cohorts: 6 diagnosed SCZ patients and 5 healthy controls. This work started with 21332 peptides of a total of 1788 proteins identified.

After the first analysis, where FDR, normalization and RSD were applied to the data, 533 peptides with a significative p-value were selected and thus, the corresponding 258 proteins were considered for this analysis. It can be recognized that most of the significant peptides, the ones that pass through the analysis, had a similar variation as the respective protein, thus, following an identical behavior or trend. However, there are some peptides that reveal a different behavior, and some examples will be shown in this chapter. One of the advantages of a peptide centered analysis is that is possible to recognize and analyze these exceptions. Thus, to achieve our goal and find what peptides can be relevant through this analysis and discovery which ones are characterized with a distinct behavior when compared to the protein and the other peptides, we apply a cut-off based on the following parameters: i) only peptides with a p-value below 0.05 could be selected, as previously referred; and ii) a value of the threshold, in module, of 50% for the fold change of each peptide vs protein fold change ratio needed to be achieved. Only the peptides and the respective proteins that pass through both parameters could be considered for the peptide centered analysis (see **supplementary table 2**).

After the variables applied and previously explained, namely FDR, normalization, RSD and the applied cut-off based on p-value and FC ratio, 37 proteins and the respective peptides, a total of 57 considered as relevant, were analyzed (see **supplementary table 2**).

5.1 Steps of analysis

We know that many peptides carry biochemical modifications and some of them can be the result of false-positive identifications. Additionally, proteins can have differential expression and within each protein we can find peptides not only with modifications but also with different area levels. To reach this information, sequential steps

were taken. In **Figure 15** can be seen the flow chart that made possible the findings shown in this chapter.

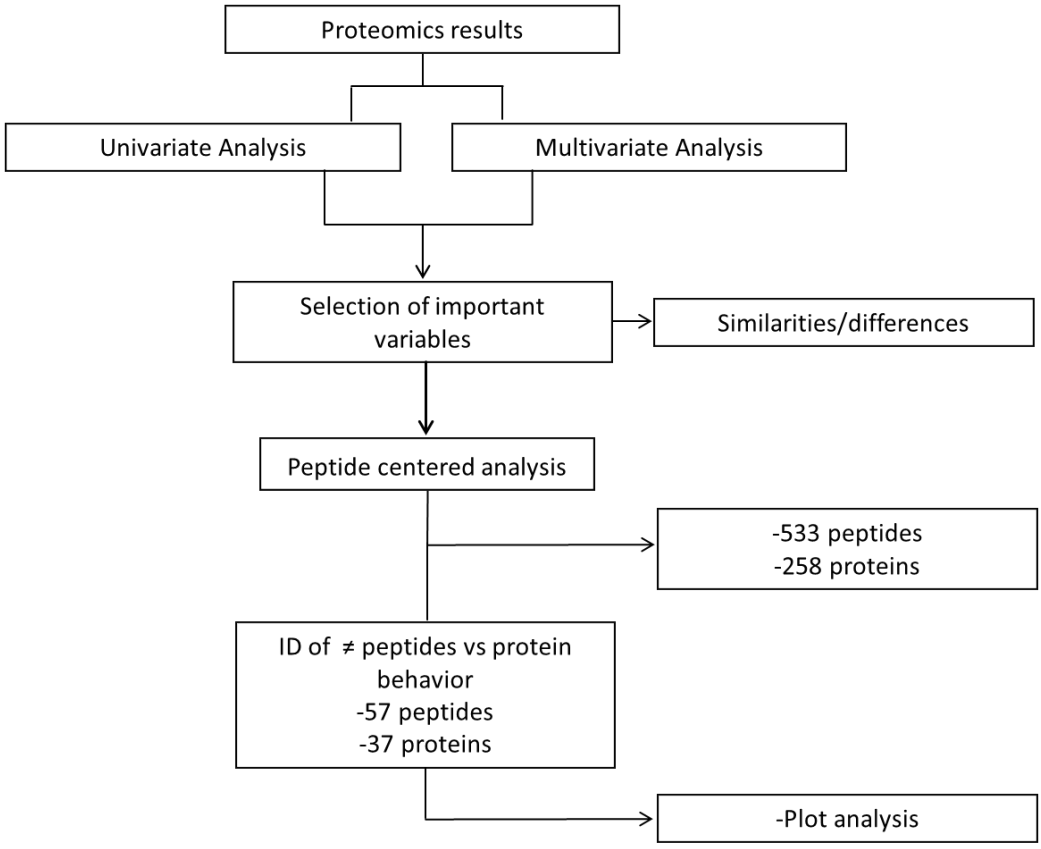


Figure 15. Flow chart of the peptides analysis process. From Proteomics results and based on the filters applied, only 533 peptides (corresponding to 258 proteins) were analyzed. These peptides were considered as relevant.

5.2 Protein vs peptides behavior

At this point, the aim was to look for the consistency of results of the proteins identified (those that pass through the univariate and multivariate analysis) and the respective peptides' behavior. Hence, the analyzes and search for differences between both were seen through a Plot result.

It could be observed that for most of the variables considered as statistically significant, there was a variation consistency of FC showed (476 peptides, which

correspond to 239 proteins). Yet, in some cases, proteins and peptides exhibited a distinct variation, which can be seen in this chapter (57 peptides, which correspond to 37 proteins).

5.2.1 sp|Q3ZCW2|LEGL_HUMAN

The first plot present here, corresponding to **Figure 16**, is a representation of the Galectin-related protein. Despite being represented only for two peptides (the ones that passed the filters applied), this protein is a good representation of what we are capable of finding in this work through a peptide-centered analysis. Additionally, a boxplot created for this data (**Figure 17**) shows the statistical significant p-values between groups expressed by asterisks in **Figure 17**. This plot helps to clarify how proteins can be modulated and masked by peptides behavior and how this information is only clear when we carry out an analysis at peptide level.

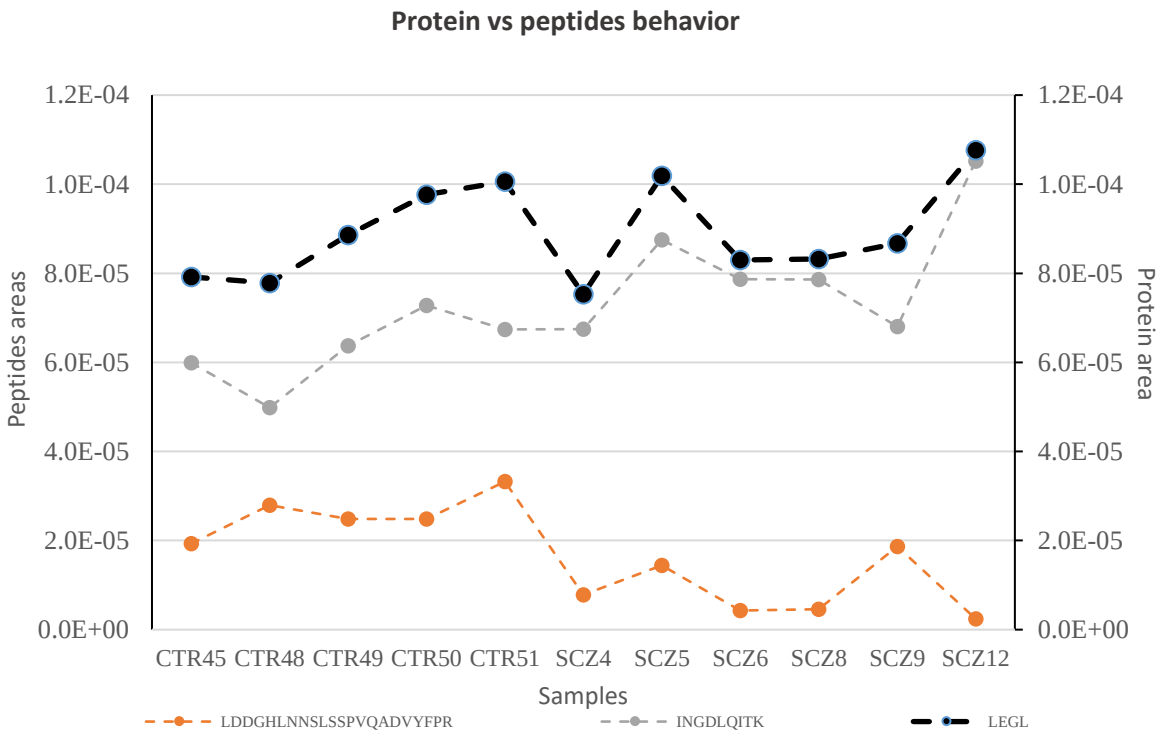


Figure 16. Protein (Galectin-related protein) vs peptide behavior. The main vertical axis shows the area values of peptides and the secondary vertical axis shows the area value of the protein. Samples are indicated on the horizontal axis. The protein analyzed is represented here by the black dashed line, while peptides have lighter and different colors. Peptides show distinct behaviors, which can be seen through the analysis done.

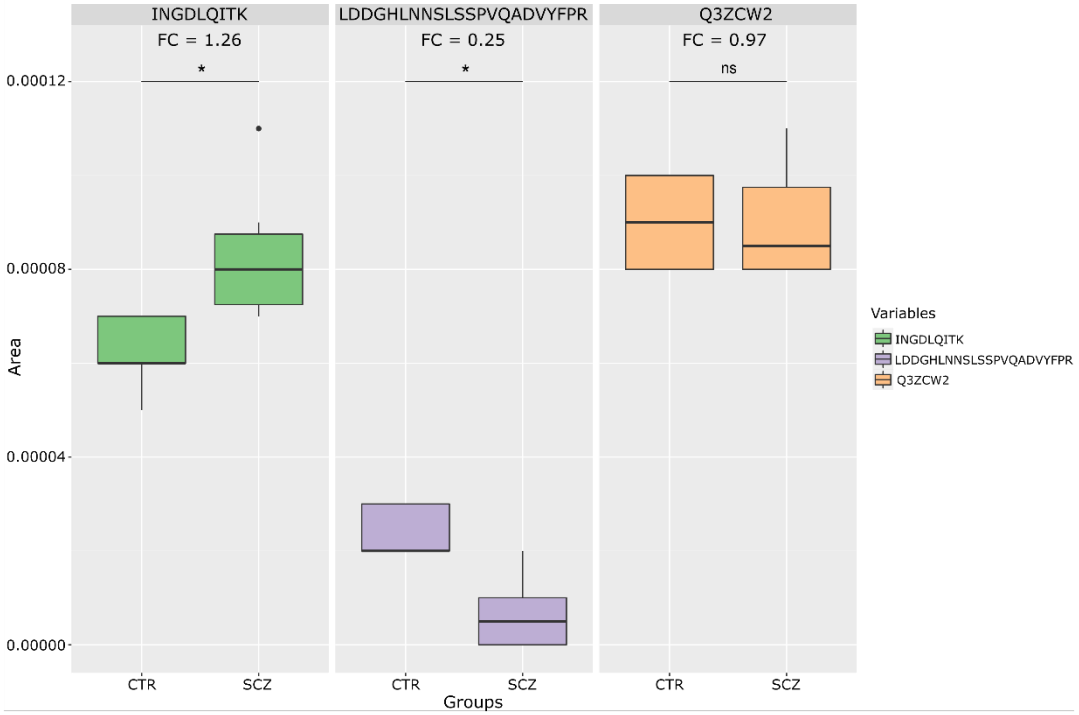


Figure 17. Boxplot of Protein (Galectin-related protein) vs peptide behavior. The vertical axis shows the area values for the peptides and the respective protein. The groups analyzed are indicated on the horizontal axis. On top of the plot is expressed the fold change (FC) values of peptides and protein. For a p-value ≤0.05, an asterisk symbol, "*", is assigned. If p-value>0.05, it is designated as not significant (ns). The variables, peptides and protein, are indicated on the right side of the plot. Looking at the groups and the area values, we can see that both peptides have a distinct behavior. However, the protein boxplot does not reflect this.

We can notice, with the support of **Figure 16** and **Figure 17**, that Galectin-related protein, as a protein, have both peptides with different behavior, especially the peptide represented by the lowest areas and orange line in the plot (**Figure 16**). The protein behavior is masked by these peptides that show to be really distinct when we compare both and look to the direction they assume.

With the support of **Figure 17**, it is even clearer that there is a distinct behavior between both peptides and within each peptide when we compare the two groups in analysis (disease and control). Analyzing both peptides, for one of them there is a statistically significant increase in the SCZ group, while the opposite occur in the other peptide, where the SCZ group have a statistically significant decrease. However, when we look at the protein, this is not evidenced. The protein does not reflect the variation trend

between the groups, which means that when we made the peptides summation, they end up annulling each other behavior.

5.2.2 sp|P02671|FIBA_HUMAN

Differently from the previously showed protein, Fibrinogen alpha chain protein (**Figure 18)** has many peptides that reach this point of analysis, and four of them are characterized for a distinct behavior. For this protein, we can see that most peptides follow the same trend, but it is also clear that some of them have a distinct behavior, which is even clear when we look for the disease group.

A recent study by Na Zhou et al. 2013, where it was also compared individuals with FEP of SCZ and healthy controls, demonstrated that there is one peptide, fibrinopeptide A (FPA), that can be a potential biomarker in the diagnosis of SCZ using MS analysis. Among the 10 fragmented peptides that showed greater discriminatory performance, they were capable of identifying the m/z fragment that corresponds to FPA. From serum samples, they did a peptide extraction before MS analysis and found that the ion at m/z 1206.6 characterized for the sequence EGDFLAEGGGVR was a fragment of fibrinopeptide A. In 2017, they used this information to analyze if the FGA gene variants that coded FPA could confer vulnerability to SCZ. However, without the expected results, their assumption was based on epigenetic regulation, which may impact the availability and function of FPA [142].

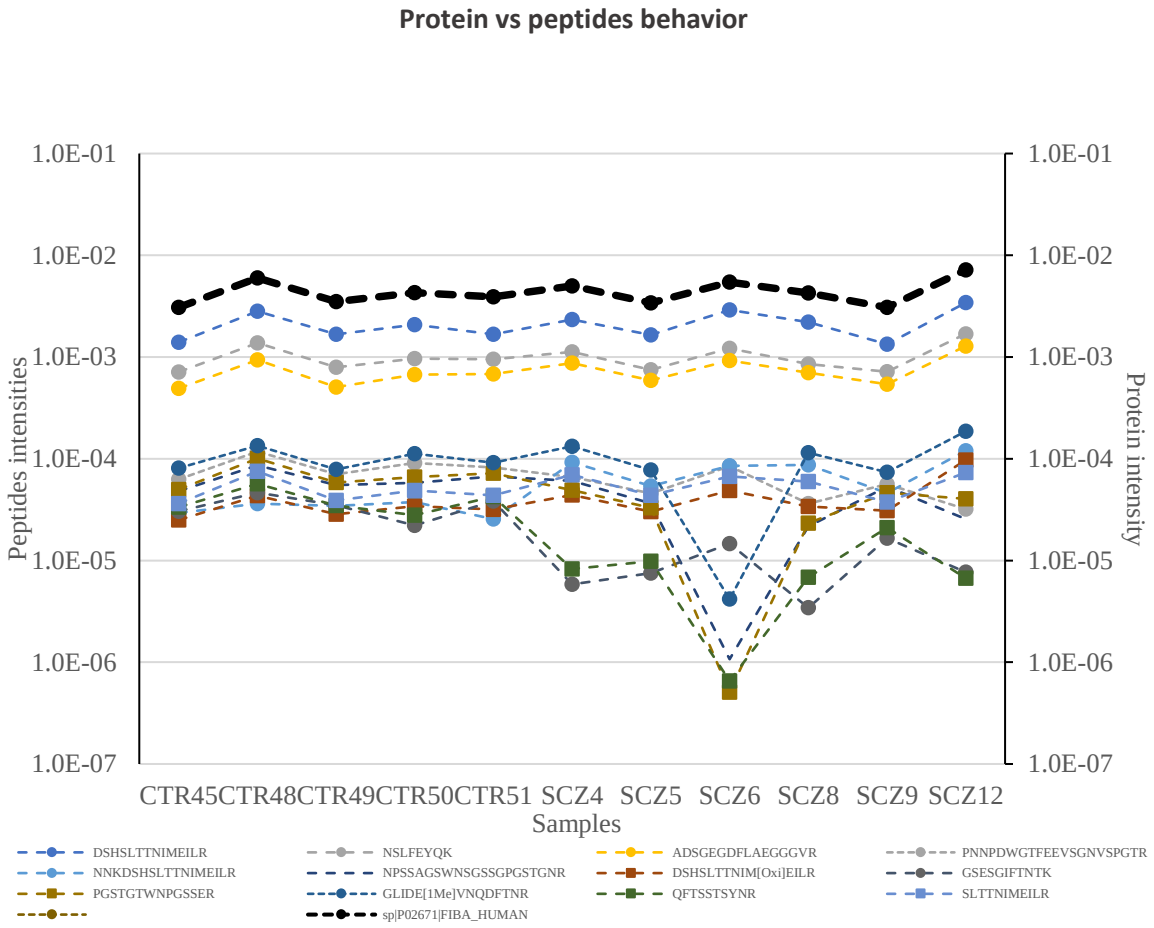


Figure 18. Protein (fibrinogen alpha chain protein) vs peptide behavior. The main vertical axis (logarithmic scale) shows the area values of peptides and the secondary vertical axis show the area value of the protein. Samples are designated in the horizontal axis. The protein analyzed is represented here by the black and dashed line, while peptides have lighter and different colors. 13 peptides were considered in this analysis and 4 of them was characterized as relevant. It is clear that most of peptides representation follows the protein line. There are 3 peptides with high area values.

These results show that some peptides have a considerable variation (more visible in **Figure 19**), and some of them can have different and notable area values between the samples in analysis, as we can see for 3 peptides with higher area in the control group, which can influence protein expression. Thus, some peptides can mask protein behavior, and if we only looked for the protein alone, it would never be perceptible. Additionally, the peptide centered analysis can create new ideas to search for potential biomarkers

and/or associations with susceptibility risks, as the study previously mentioned reported and aimed to investigate, respectively.

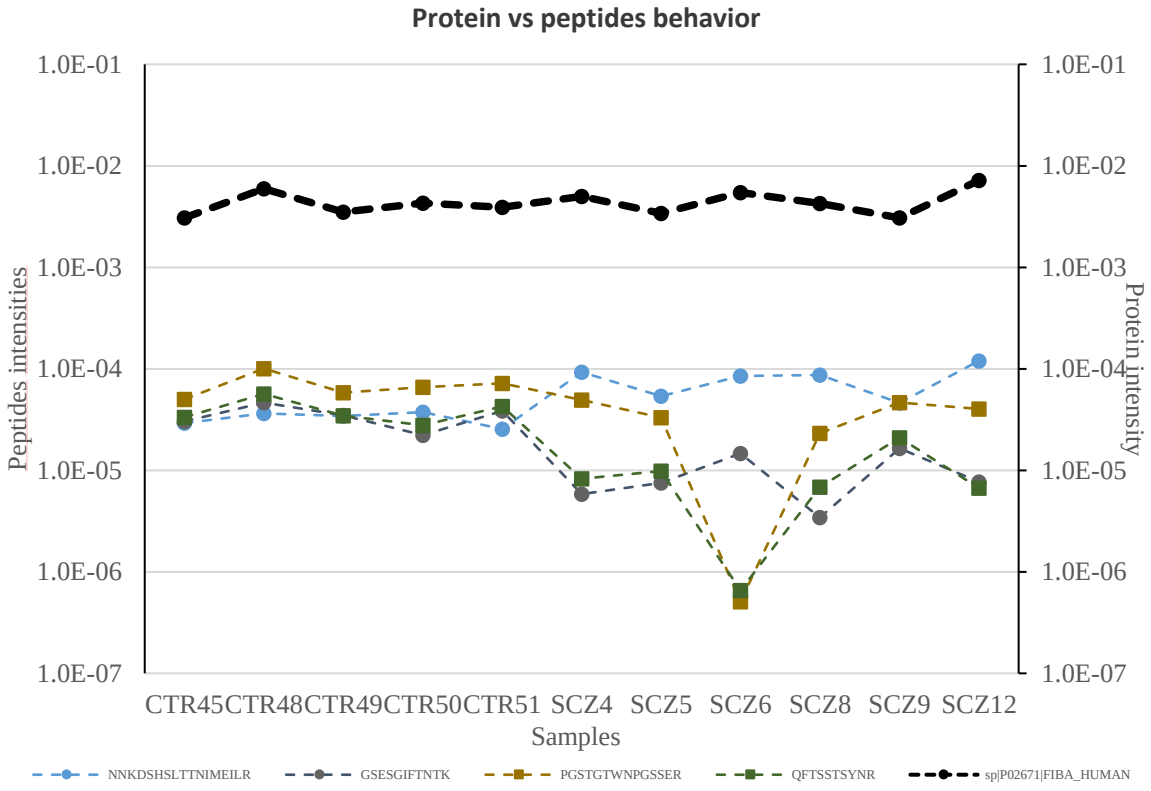


Figure 19. Protein (fibrinogen alpha chain protein) vs peptide behavior. The main vertical axis (logarithmic scale) shows the area values of peptides and the secondary vertical axis show the area value of the protein. Samples are designated in the horizontal axis. The protein analyzed is represented here by the black and dashed line, while peptides have lighter and different colors. In this figure, only the relevant peptides, 4 in a total of 13, are shown.

5.2.3 sp|P07437|TBB5_HUMAN

Another protein where we can see a distinct peptide behavior is the Tubulin beta chain. This protein, coded by the TUBB gene, has an isoform (Tubulin β-III) that is known as the most dynamic and is highly expressed in CNS [143]. In a study of Rodrigues-Amorim, D. et al. 2020, a comparative study between patients with SCZ (FEP and chronic) and healthy subjects was performed in order to evaluate the potential neuronal damage. Thus, they analyzed structural proteins, one of them Tubulin β-III. Different from our study, they used plasma samples and immunoblotting to analyze the protein levels. Significant differences

between patients and controls were found for this protein. Additionally, patients who did not respond to the first-line atypical Aps, presented higher levels of Tubulin β-III. Thus, they conclude that besides this protein may predict the neurodegenerative course of SCZ; it could be able to predict the progression of the disease.

In our results, one of the three peptides that characterize this protein shows a different behavior and has different area levels (**Figure 20**). Despite some differences in area levels, the other two peptides have similar behavior and explain the protein trend. Once more, some peptides can mask protein behavior, and only through this type of analysis it is possible to recognize that since proteomic analysis is focused on proteins total area. More detailed information can be achieved when this type of analysis is performed and, as shown in the study of Na Zhou et al. 2013, it can be the starting point for future research.

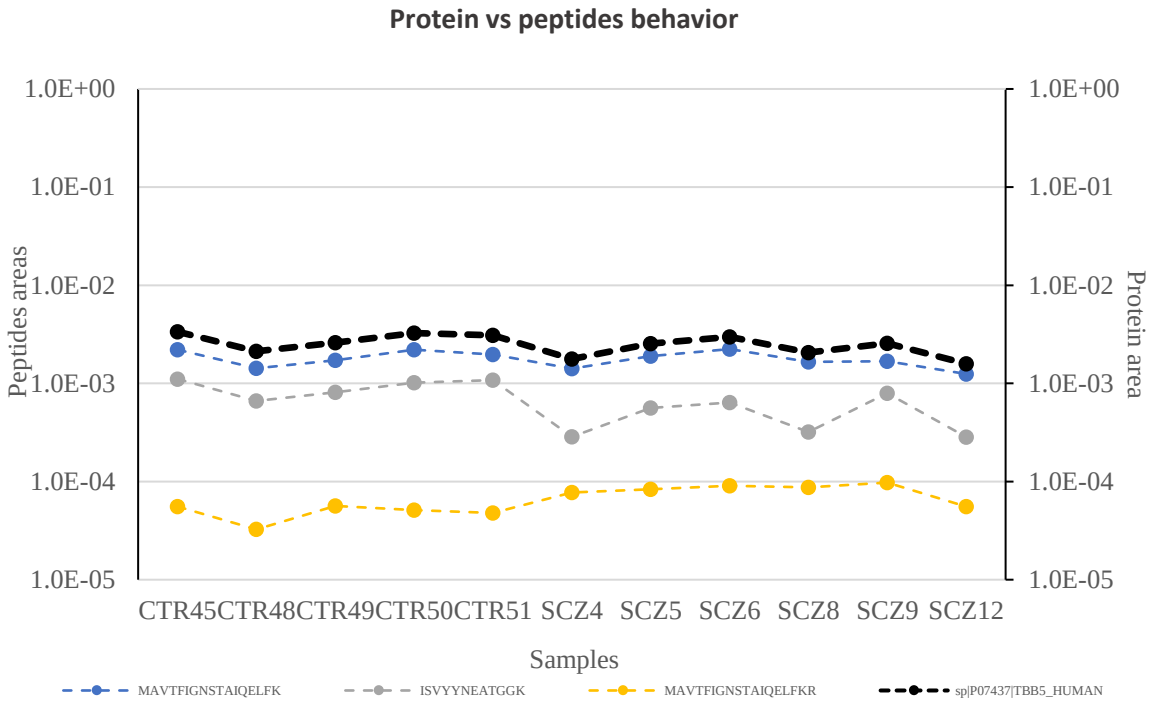


Figure 20. Protein (tubulin beta chain) vs peptide behavior. The main vertical axis shows the area values of peptides and the secondary vertical axis shows the area value of the protein. Samples are indicated on the horizontal axis. The protein analyzed is represented here by the black and dashed line, while peptides have lighter and different colors. In this figure, there is only a relevant peptide, among three peptides, which is here represented by the yellow line.

This analysis shows that different approaches can be taken into account when we analyze this type of data. When this analysis was started, it was first recognized that there are distinct proteins between both groups, disease and control. Usually, the focus is the proteins, but if we take a look at the peptide level, there is some information that can lead us to news studies.

Peptides with distinct behavior, as it was shown in some studies already mentioned here, create the need to explore and understand why these peptides stand out among the others and how the respective protein can be associated with the disease, exploring, for example, possible variants of the gene that coded that protein. What we can reach through this analysis is that most of the peptides identified as relevant for group separation are in agreement with protein results, i.e., they follow the same direction as the respective protein. However, there is a specific group of peptides with distinct behavior, and these results can also explain how some modifications may be a product of PTMs. If, on the one hand, they can diversify protein functions, being essential for the cells, one the other hand, perturbations of PTMs in cells can be prejudicial for the maintenance of normal cellular states. Moreover, the development of disorders and human diseases have been linked to PTMs defects [144]. However, this can be a result of different proteoform types as alternative splicing of a pre-mRNA [145]. Thus, based on our results, two other approaches can be reported, and they are centered on peptides: i) an approach where peptides are different between the groups in analysis, and ii) an approach where some peptides have different behavior when compared to the protein. With the focus on the last, as it has been said, some protein isoforms can explain the behavior and some diseases. One well-known case is Alzheimer's disease, a neurodegenerative disease where the protein Microtubule-associated protein Tau can undergo distinct post-translational modifications. The hyperphosphorylated tau protein is a biological marker that characterizes the disease and leads to protein aggregation in neurons through brain lesions as neurofibrillary tangles (NFTs) [146, 147]. This knowledge has supported and conducted extensive studies until now and shows how this kind of discovery and information can be useful for a better understanding of the pathophysiology of diseases and for biomarker discovery. Thus, a different peptide behavior can make us think in PTMs or another type of proteoform and can be the starting point for future studies, as to understand why some PTMs happen between both groups. Additionally, if we exclude these peptides and only consider the peptides with the same behavior of the respective protein, we can use those sequences to apply to monitoring techniques. Based on that selection, antibodies can be generated from specific peptides and being used in enzyme-linked immunosorbent assay (ELISA), for example, or an MS target approach can be selected to monitor those specific peptides.

CHAPTER 6- Conclusions

6 Conclusions

Biomarkers discovery can hopefully improve the diagnosis and the monitoring of diseases. Psychiatric disorders, namely Schizophrenia, lack diagnostic tests. It has been reported that the time from the beginning of the disease and the correct treatment is crucial for the prognostic of the disease. As a disease that occurs early in life and compromises the life of the individuals affected, it is imperative the need to find biomarkers that can contribute to patient's health.

During this work, a proteomic analysis was performed, followed by a Multivariate Analysis. As a result, we could notice a clear separation between the groups in analysis, with this difference being explained by the identification of 396 proteins and 1338 peptides with VIP higher than one. Among these results, achieved through a PLS-DA model, 118 proteins and 533 peptides were characterized with a p-value <0.05. Considering the proteins with statistical meaning, Gene Ontology analysis provided information of the pathways altered and identified the enriched ones. The pathways characterized as altered are linked to hemostasis function with association to platelet function and coagulation cascade. The results, which are in concordance with other studies performed in individuals with SCZ, show that the approach applied allows identifying biological alterations related to the disease.

Additionally, it was presented a different and unusual strategy to evaluate and search for differences between a disease group (SCZ patients) and a group of control (healthy individuals). This type of analysis is capable of showing what in a conventional analysis is not noticeable, as a different peptide and protein behavior. Consequently, it can be the starting point for new studies, as promoting studies of specific PTMs or for a specific protein, or even, in a better way, lead us to make the selection of some of the peptides that can be further used in larger cohorts using targeted approaches (MRM or ELISA) to stratify individuals better.

Through the analysis performed, although most peptides show that they have the same trend as the protein, a considerable number of peptides have distinct behavior. In this analysis, we were able to recognize that 57 peptides have a different behavior compared to the respective 37 proteins.

Since a protein-centered analysis is focused on total area and as some of these trends can only be recognized through a peptide-centered analysis, it could be interesting to do more research and explore these results in future studies. Moreover, it could help or even be a possible approach in the search for biomarkers.

We know that proteomics approaches provides qualitative and quantitative information about proteins. However, these quantitation strategies presume that protein peptides will have an identical quantitative behavior. This assumption can be wrong and this analysis is demonstrating that even within the same group, we can have the presence of heterogeneous proteoforms, which is seen by different peptides behavior. Through MS analysis, we can analyze those differences and this can be useful to the comprehensive understanding of biological system. Additionally, the knowledge of proteoforms and proteoform families, with the report of this kind of information into a specific database, will provide a fast and helpful identification of proteoforms in the future.

CHAPTER 7- References

7 References

1. Hyman, S.E., *A glimmer of light for neuropsychiatric disorders.* Nature, 2008. **455**(7215): p. 890-3.
2. Peedicayil, J., *Identification of Biomarkers in Neuropsychiatric Disorders Based on Systems Biology and Epigenetics.* Front Genet, 2019. **10**: p. 985.
3. Guest, F.L., P.C. Guest, and D. Martins-de-Souza, *The emergence of point-of-care blood-based biomarker testing for psychiatric disorders: enabling personalized medicine.* Biomark Med, 2016. **10**(4): p. 431-43.
4. Taurines, R., et al., *Proteomic research in psychiatry.* J Psychopharmacol, 2011. **25**(2): p. 151-96.
5. Nestler, E.J. and S.E. Hyman, *Animal models of neuropsychiatric disorders.* Nat Neurosci, 2010. **13**(10): p. 1161-9.
6. Comes, A.L., et al., *Proteomics for blood biomarker exploration of severe mental illness: pitfalls of the past and potential for the future.* Transl Psychiatry, 2018. **8**(1): p. 160.
7. Trautmann, S., J. Rehm, and H.U. Wittchen, *The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders?* EMBO Rep, 2016. **17**(9): p. 1245-9.
8. Organization, W.H., *World report on disability 2011*. 2011: World Health Organization.
9. Preece, R.L., S.Y.S. Han, and S. Bahn, *Proteomic approaches to identify blood-based biomarkers for depression and bipolar disorders.* Expert Review of Proteomics, 2018. **15**(4): p. 325-340.
10. Zuccoli, G.S., et al., *The Energy Metabolism Dysfunction in Psychiatric Disorders Postmortem Brains: Focus on Proteomic Evidence.* Frontiers in Neuroscience, 2017. **11**: p. 14.
11. Taber, K.H., R.A. Hurley, and S.C. Yudofsky, *Diagnosis and treatment of neuropsychiatric disorders.* Annu Rev Med, 2010. **61**: p. 121-33.
12. Vawter, M.P., et al., *Exon Array Biomarkers for the Differential Diagnosis of Schizophrenia and Bipolar Disorder.* Mol Neuropsychiatry, 2018. **3**(4): p. 197-213.
13. Lago, S.G., et al., *Exploring the neuropsychiatric spectrum using high-content functional analysis of single-cell signaling networks.* Mol Psychiatry, 2018.
14. Ayyagari, R., et al., *Association of antipsychotic treatment switching in patients with schizophrenia, bipolar, and major depressive disorders.* J Med Econ, 2020. **23**(2): p. 204-212.
15. Santa, C., et al., *Circulating biomarkers in schizophrenia: a proteomics perspective.* International Journal of Clinical Neurosciences and Mental Health, 2017(4(Suppl. 3)): p. S05.
16. Gustavsson, A., et al., *Cost of disorders of the brain in Europe 2010.* European neuropsychopharmacology, 2011. **21**(10): p. 718-779.
17. Lakhan, S.E. and A. Kramer, *Schizophrenia genomics and proteomics: are we any closer to biomarker discovery?* Behav Brain Funct, 2009. **5**: p. 2.
18. Nascimento, J.M. and D. Martins-de-Souza, *The proteome of schizophrenia.* NPJ Schizophr, 2015. **1**: p. 14003.
19. Yin, L., et al., *Leveraging genome-wide association and clinical data in revealing schizophrenia subgroups.* J Psychiatr Res, 2018. **106**: p. 106-117.
20. Charlson, F.J., et al., *Global Epidemiology and Burden of Schizophrenia: Findings From the Global Burden of Disease Study 2016.* Schizophr Bull, 2018. **44**(6): p. 1195-1203.
21. Schwarz, E. and S. Bahn, *Biomarker discovery in psychiatric disorders.* Electrophoresis, 2008. **29**(13): p. 2884-90.

22. Giusti, L., et al., *Proteomics insight into psychiatric disorders: an update on biological fluid biomarkers.* Expert Rev Proteomics, 2016. **13**(10): p. 941-950.
23. Alawam, K., *Application of proteomics in diagnosis of ADHD, schizophrenia, major depression, and suicidal behavior.* Adv Protein Chem Struct Biol, 2014. **95**: p. 283-315.
24. Tomasik, J., et al., *Neuroimmune biomarkers in schizophrenia.* Schizophr Res, 2016. **176**(1): p. 3-13.
25. Cohen, C.I., P.D. Meesters, and J. Zhao, *New perspectives on schizophrenia in later life: implications for treatment, policy, and research.* The Lancet Psychiatry, 2015. **2**(4): p. 340-350.
26. Aricioglu, F., et al., *Neuroinflammation in Schizophrenia: A Critical Review and The Future.* Klinik Psikofarmakoloji Bülteni-Bulletin of Clinical Psychopharmacology, 2017. **26**(4): p. 429-437.
27. Davison, J., et al., *A systematic review of metabolite biomarkers of schizophrenia.* Schizophr Res, 2018. **195**: p. 32-50.
28. Arciniegas, D.B., *Psychosis.* Continuum (Minneap Minn), 2015. **21**(3 Behavioral Neurology and Neuropsychiatry): p. 715-36.
29. McGorry, P.D., et al., *Back to the future: predicting and reshaping the course of psychotic disorder.* Archives of General Psychiatry, 2008. **65**(1): p. 25-27.
30. Owen, M.J., A. Sawa, and P.B. Mortensen, *Schizophrenia.* The Lancet, 2016. **388**(10039): p. 86-97.
31. Kumar, A., et al., *Potential drug targets and treatment of schizophrenia.* Inflammopharmacology, 2017. **25**(3): p. 277-292.
32. Lally, J. and J.H. MacCabe, *Antipsychotic medication in schizophrenia: a review.* Br Med Bull, 2015. **114**(1): p. 169-79.
33. Yang, A.C. and S.J. Tsai, *New Targets for Schizophrenia Treatment beyond the Dopamine Hypothesis.* Int J Mol Sci, 2017. **18**(8).
34. Howes, O., R. McCutcheon, and J. Stone, *Glutamate and dopamine in schizophrenia: an update for the 21st century.* J Psychopharmacol, 2015. **29**(2): p. 97-115.
35. Miyamoto, S., et al., *Treatments for schizophrenia: a critical review of pharmacology and mechanisms of action of antipsychotic drugs.* Mol Psychiatry, 2005. **10**(1): p. 79-104.
36. Corigliano, V., et al., *Long-Acting Injectable Second-Generation Antipsychotics Improve Negative Symptoms and Suicidal Ideation in Recent Diagnosed Schizophrenia Patients: A 1-Year Follow-up Pilot Study.* Schizophr Res Treatment, 2018. **2018**: p. 4834135.
37. Kane, J.M. and C.U. Correll, *Optimizing Treatment Choices to Improve Adherence and Outcomes in Schizophrenia.* The Journal of clinical psychiatry, 2019. **80**(5).
38. Correll, C.U., et al., *The use of long-acting injectable antipsychotics in schizophrenia: evaluating the evidence.* J Clin Psychiatry, 2016. **77**(suppl 3): p. 1-24.
39. K Shin, J., et al., *Schizophrenia: a systematic review of the disease state, current therapeutics and their molecular mechanisms of action.* Current medicinal chemistry, 2011. **18**(9): p. 1380-1404.
40. Christian, R., et al., *Future research needs for first-and second-generation antipsychotics for children and young adults.* 2012.
41. Frankel, J.S. and T.L. Schwartz, *Brexpiprazole and cariprazine: distinguishing two new atypical antipsychotics from the original dopamine stabilizer aripiprazole.* Ther Adv Psychopharmacol, 2017. **7**(1): p. 29-41.
42. Shin, J.K., et al., *Schizophrenia: a systematic review of the disease state, current therapeutics and their molecular mechanisms of action.* Curr Med Chem, 2011. **18**(9): p. 1380-404.

43. Guest, P.C., et al., *Abnormalities in metabolism and hypothalamic-pituitary-adrenal axis function in schizophrenia.* Int Rev Neurobiol, 2011. **101**: p. 145-68.
44. Tzeng, W.C., et al., *Gender differences in metabolic syndrome risk factors among patients with serious mental illness.* Int J Ment Health Nurs, 2019.
45. Pantelis, C., et al., *Structural brain imaging evidence for multiple pathological processes at different stages of brain development in schizophrenia.* Schizophr Bull, 2005. **31**(3): p. 672-96.
46. Davalieva, K., I.M. Kostovska, and A.J. Dwork, *Proteomics Research in Schizophrenia.* Frontiers in Cellular Neuroscience, 2016. **10**: p. 22.
47. Kahn, R.S. and I.E. Sommer, *The neurobiology and treatment of first-episode schizophrenia.* Mol Psychiatry, 2015. **20**(1): p. 84-97.
48. Patel, S., *Role of proteomics in biomarker discovery: prognosis and diagnosis of neuropsychiatric disorders.* Adv Protein Chem Struct Biol, 2014. **94**: p. 39-75.
49. Bradley, A.J. and T.G. Dinan, *A systematic review of hypothalamic-pituitary-adrenal axis function in schizophrenia: implications for mortality.* J Psychopharmacol, 2010. **24**(4 Suppl): p. 91-118.
50. Müller, N., *Inflammation in schizophrenia: pathogenetic aspects and therapeutic considerations.* Schizophrenia bulletin, 2018. **44**(5): p. 973-982.
51. Girgis, R.R., S.S. Kumar, and A.S. Brown, *The cytokine model of schizophrenia: emerging therapeutic strategies.* Biol Psychiatry, 2014. **75**(4): p. 292-299.
52. Brisch, R., et al., *The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: old fashioned, but still in vogue.* Front Psychiatry, 2014. **5**: p. 47.
53. Sayin, H., *Getting high on dopamine: Neuro scientific aspects of pleasure.* SexuS Journal, 2019. **4**(11): p. 883-906.
54. Snyder, M.A. and W.J. Gao, *NMDA hypofunction as a convergence point for progression and symptoms of schizophrenia.* Front Cell Neurosci, 2013. **7**: p. 31.
55. Kesby, J., et al., *Dopamine, psychosis and schizophrenia: the widening gap between basic and clinical neuroscience.* Translational psychiatry, 2018. **8**(1): p. 1-12.
56. Lieberman, J.A., J.M. Kane, and J. Alvir, *Provocative tests with psychostimulant drugs in schizophrenia.* Psychopharmacology (Berl), 1987. **91**(4): p. 415-33.
57. Howes, O.D. and S. Kapur, *The dopamine hypothesis of schizophrenia: version III--the final common pathway.* Schizophr Bull, 2009. **35**(3): p. 549-62.
58. Seeman, P. and S. Kapur, *Schizophrenia: more dopamine, more D2 receptors.* Proc Natl Acad Sci U S A, 2000. **97**(14): p. 7673-5.
59. Schizophrenia Working Group of the Psychiatric Genomics, C., *Biological insights from 108 schizophrenia-associated genetic loci.* Nature, 2014. **511**(7510): p. 421-7.
60. Zargar, F., et al., *Effectiveness of Dialectical Behavioral Therapy on Executive Function, Emotional Control and Severity of Symptoms in Patients with Bipolar I Disorder.* Adv Biomed Res, 2019. **8**: p. 59.
61. Smith, D.J., E.A. Whitham, and S.N. Ghaemi, *Bipolar disorder.* Handb Clin Neurol, 2012. **106**: p. 251-63.
62. Lin, K., et al., *Inflammation, brain structure and cognition interrelations among individuals with differential risks for bipolar disorder.* Brain Behav Immun, 2020. **83**: p. 192-199.
63. Roda, A., I. Chendo, and M. Kunz, *Biomarkers and staging of bipolar disorder: a systematic review.* Trends Psychiatry Psychother, 2015. **37**(1): p. 3-11.
64. Fries, G.R., et al., *Hypothalamic-pituitary-adrenal axis dysfunction and illness progression in bipolar disorder.* Int J Neuropsychopharmacol, 2014. **18**(1).

65. Roshanaei-Moghaddam, B. and W. Katon, *Premature mortality from general medical illnesses among persons with bipolar disorder: a review.* Psychiatr Serv, 2009. **60**(2): p. 147-56.
66. Cheng, Y., et al., *Elevated heat shock proteins in bipolar disorder patients with hypothalamic pituitary adrenal axis dysfunction.* Medicine (Baltimore), 2018. **97**(27): p. e11089.
67. Wang, S.S., et al., *Gene expression analysis in the human hypothalamus in depression by laser microdissection and real-time PCR: the presence of multiple receptor imbalances.* Mol Psychiatry, 2008. **13**(8): p. 786-99, 741.
68. van Beveren, N.J. and W.J. Hoogendijk, *Clinical utility of serum biomarkers for major psychiatric disorders.* Int Rev Neurobiol, 2011. **101**: p. 351-74.
69. Paris, J. and D.W. Black, *Borderline personality disorder and bipolar disorder: what is the difference and why does it matter?* J Nerv Ment Dis, 2015. **203**(1): p. 3-7.
70. Focking, M., et al., *Proteomic analysis of the postsynaptic density implicates synaptic function and energy pathways in bipolar disorder.* Transl Psychiatry, 2016. **6**(11): p. e959.
71. Hill, S.K., et al., *Neurocognitive allied phenotypes for schizophrenia and bipolar disorder.* Schizophr Bull, 2008. **34**(4): p. 743-59.
72. Filiou, M.D. and C.W. Turck, *General overview: biomarkers in neuroscience research.* Int Rev Neurobiol, 2011. **101**: p. 1-17.
73. Canuel, V., et al., *Translational research platforms integrating clinical and omics data: a review of publicly available solutions.* Brief Bioinform, 2015. **16**(2): p. 280-90.
74. McShane, L.M., et al., *Criteria for the use of omics-based predictors in clinical trials.* Nature, 2013. **502**(7471): p. 317-20.
75. Biomarkers Definitions Working, G., *Biomarkers and surrogate endpoints: preferred definitions and conceptual framework.* Clin Pharmacol Ther, 2001. **69**(3): p. 89-95.
76. Guest, P.C., et al., *The use of proteomic biomarkers for improved diagnosis and stratification of schizophrenia patients.* Biomark Med, 2014. **8**(1): p. 15-27.
77. Bayes, A. and S.G. Grant, *Neuroproteomics: understanding the molecular organization and complexity of the brain.* Nat Rev Neurosci, 2009. **10**(9): p. 635-46.
78. Kitchen, R.R., et al., *Decoding neuroproteomics: integrating the genome, translatome and functional anatomy.* Nat Neurosci, 2014. **17**(11): p. 1491-9.
79. Santa, C., et al., *Neuroproteomics — LC-MS Quantitative Approaches.* 2015.
80. Gottschalk, M.G., et al., *Proteomic enrichment analysis of psychotic and affective disorders reveals common signatures in presynaptic glutamatergic signaling and energy metabolism.* Int J Neuropsychopharmacol, 2014. **18**(2).
81. Sabherwal, S., et al., *Blood biomarker discovery in drug-free schizophrenia: the contributionof proteomics and multiplex immunoassays.* Expert Rev Proteomics, 2016. **13**(12): p. 1141-1155.
82. Lai, C.Y., et al., *Biomarkers in schizophrenia: A focus on blood based diagnostics and theranostics.* World J Psychiatry, 2016. **6**(1): p. 102-17.
83. de Jesus, J.R., et al., *Bipolar disorder: recent advances and future trends in bioanalytical developments for biomarker discovery.* Anal Bioanal Chem, 2015. **407**(3): p. 661-7.
84. Chan, M.K., et al., *Converging evidence of blood-based biomarkers for schizophrenia: an update.* Int Rev Neurobiol, 2011. **101**: p. 95-144.
85. Giusti, L., et al., *Search for peripheral biomarkers in patients affected by acutely psychotic bipolar disorder: a proteomic approach.* Mol Biosyst, 2014. **10**(6): p. 1246-54.
86. Nascimento, J.M., et al., *Proteomics and molecular tools for unveiling missing links in the biochemical understanding of schizophrenia.* Proteomics Clin Appl, 2016. **10**(12): p. 1148-1158.

87. Ginsburg, G.S. and J.J. McCarthy, *Personalized medicine: revolutionizing drug discovery and patient care.* TRENDS in Biotechnology, 2001. **19**(12): p. 491-496.
88. Quattrocchi, A., et al., *Personalized medicine in psychiatric disorders: prevention and bioethical questions.* La Clinica Terapeutica, 2019. **170**(6): p. e421-e424.
89. Nice, E.C., *From proteomics to personalized medicine: the road ahead.* Expert Rev Proteomics, 2016. **13**(4): p. 341-3.
90. Personalized Medicine Coalition. *Personalized Medicine at FDA: The Scope & Significance of Progress in 2019*. 2020, August 21; Available from: http://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/PM_at_FDA_The_Scope_and_Significance_of_Progress_in_2019.pdf.
91. Llerena, A., et al., *Low frequency of CYP2D6 poor metabolizers among schizophrenia patients.* The pharmacogenomics journal, 2007. **7**(6): p. 408-410.
92. de Jesus, J.R., et al., *Bipolar disorder: recent advances and future trends in bioanalytical developments for biomarker discovery.* Analytical and Bioanalytical Chemistry, 2015. **407**(3): p. 661-667.
93. Guest, P.C., F.L. Guest, and D. Martins-de Souza, *Making Sense of Blood-Based Proteomics and Metabolomics in Psychiatric Research.* Int J Neuropsychopharmacol, 2016. **19**(6).
94. Wei, X. and L. Li, *Mass spectrometry-based proteomics and peptidomics for biomarker discovery in neurodegenerative diseases.* Int J Clin Exp Pathol, 2009. **2**(2): p. 132-48.
95. Cunningham, R., D. Ma, and L. Li, *Mass Spectrometry-based Proteomics and Peptidomics for Systems Biology and Biomarker Discovery.* Front Biol (Beijing), 2012. **7**(4): p. 313-335.
96. Tuli, L. and H.W. Ressom, *LC-MS Based Detection of Differential Protein Expression.* J Proteomics Bioinform, 2009. **2**: p. 416-438.
97. Aebersold, R. and M. Mann, *Mass-spectrometric exploration of proteome structure and function.* Nature, 2016. **537**(7620): p. 347-55.
98. Chandramouli, K. and P.Y. Qian, *Proteomics: challenges, techniques and possibilities to overcome biological sample complexity.* Hum Genomics Proteomics, 2009. **2009**.
99. de Hoffmann, E. and V. Stroobant, *Mass Spectrometry: Principles and Applications*. 2007: Wiley.
100. Di Girolamo, F., et al., *The Role of Mass Spectrometry in the "Omics" Era.* Curr Org Chem, 2013. **17**(23): p. 2891-2905.
101. Pitt, J.J., *Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry.* Clin Biochem Rev, 2009. **30**(1): p. 19-34.
102. Kinter, M. and N.E. Sherman, *Protein sequencing and identification using tandem mass spectrometry*. Vol. 9. 2005: John Wiley & Sons.
103. Medhe, S., *Mass spectrometry: Detectors review.* Chemical and Biomolecular Engineering, 2018. **3**(4): p. 51.
104. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review.* Anal Bioanal Chem, 2007. **389**(4): p. 1017-31.
105. Walker, S.J. and A. Xu, *Biomarker discovery using molecular profiling approaches*, in *International review of neurobiology*. 2004, Elsevier. p. 1-30.
106. Neilson, K.A., et al., *Less label, more free: approaches in label-free quantitative mass spectrometry.* Proteomics, 2011. **11**(4): p. 535-53.
107. Gillet, L.C., A. Leitner, and R. Aebersold, *Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing.* Annu Rev Anal Chem (Palo Alto Calif), 2016. **9**(1): p. 449-72.
108. Anand, S., et al., *Label-Based and Label-Free Strategies for Protein Quantitation.* Methods Mol Biol, 2017. **1549**: p. 31-43.

109. Anjo, S.I., C. Santa, and B. Manadas, *SWATH-MS as a tool for biomarker discovery: From basic research to clinical applications.* Proteomics, 2017. **17**(3-4).
110. Ludwig, C., et al., *Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial.* Mol Syst Biol, 2018. **14**(8): p. e8126.
111. Gillet, L.C., et al., *Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis.* Mol Cell Proteomics, 2012. **11**(6): p. O111 016717.
112. Özdemir, V., et al., *Personalized medicine beyond genomics: alternative futures in big data—proteomics, environtome and the social proteome.* Journal of Neural Transmission, 2017. **124**(1): p. 25-32.
113. Sanchez-Pinto, L.N., Y. Luo, and M.M. Churpek, *Big data and data science in critical care.* Chest, 2018. **154**(5): p. 1239-1248.
114. Popescu, G.V., C. Noutsos, and S.C. Popescu, *Big data in plant science: resources and data mining tools for plant genomics and proteomics*, in *Data Mining Techniques for the Life Sciences*. 2016, Springer. p. 533-547.
115. Bensadek, D., A. Nicolas, and A.I. Lamond, *Next Generation Proteomics: PTMs in Space and Time.* Quantitative proteomics, 2014(1): p. 233.
116. Wilhelm, M., et al., *Mass-spectrometry-based draft of the human proteome.* Nature, 2014. **509**(7502): p. 582-7.
117. Karimpour, M., *Multi-platform metabolomics assays to study the responsiveness of the human plasma and lung lavage metabolome*. 2016, Umeå University.
118. Stein, S.A.M., et al., *Principal Components Analysis: A Review of its Application on Molecular Dynamics Data.* Annual reports in computational chemistry. 2, 2005. **1**: p. 233.
119. Abdi, H., *Partial least squares regression and projection on latent structure regression (PLS Regression).* Wiley interdisciplinary reviews: computational statistics, 2010. **2**(1): p. 97-106.
120. Anjo, S.I., et al., *Unraveling Mesenchymal Stem Cells' Dynamic Secretome Through Nontargeted Proteomics Profiling*, in *Methods Mol Biol*. 2016. p. 521-49.
121. Anjo, S.I., et al., *Use of recombinant proteins as a simple and robust normalization method for untargeted proteomics screening: exhaustive performance assessment.* Talanta, 2019. **205**: p. 120163.
122. Anjo, S.I., C. Santa, and B. Manadas, *Short GeLC-SWATH: a fast and reliable quantitative approach for proteomic screenings.* Proteomics, 2015. **15**(4): p. 757-62.
123. Manadas, B., et al., *BDNF-induced changes in the expression of the translation machinery in hippocampal neurons: protein levels and dendritic mRNA.* J Proteome Res, 2009. **8**(10): p. 4536-52.
124. Anjo, S.I., et al., *Neuroproteomics using short GeLC-SWATH: from the evaluation of proteome changes to the clarification of protein function*, in *Current Proteomic Approaches Applied to Brain Function*. 2017, Springer. p. 107-138.
125. Sennels, L., J.-C. Bukowski-Wills, and J. Rappsilber, *Improved results in proteomics by use of local and peptide-class specific false discovery rates.* BMC bioinformatics, 2009. **10**: p. 179-179.
126. Tang, W.H., I.V. Shilov, and S.L. Seymour, *Nonlinear fitting method for determining local false discovery rates from decoy database searches.* Journal of proteome research, 2008. **7**(9): p. 3661-7.
127. Oliveira, A.I., et al., *Crosstalk between glial and glioblastoma cells triggers the "go-or-grow" phenotype of tumor cells.* Cell Communication and Signaling, 2017. **15**(1): p. 1-12.
128. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Journal of the Royal statistical society: series B (Methodological), 1995. **57**(1): p. 289-300.

129. Bogdanow, B., H. Zauber, and M. Selbach, *Systematic Errors in Peptide and Protein Identification and Quantification by Modified Peptides.* Mol Cell Proteomics, 2016. **15**(8): p. 2791-801.
130. Valikangas, T., T. Suomi, and L.L. Elo, *A systematic evaluation of normalization methods in quantitative label-free proteomics.* Brief Bioinform, 2018. **19**(1): p. 1-11.
131. Alzate, O., *Neuroproteomics*. 2009: CRC Press.
132. Tanizaki, H., *Power comparison of non-parametric tests: Small-sample properties from Monte Carlo experiments.* Journal of applied statistics, 1997. **24**(5): p. 603-632.
133. Lo, S., et al., *Non-significant in univariate but significant in multivariate analysis: a discussion with examples.* Changgeng Yi Xue Za Zhi, 1995. **18**(2): p. 95-101.
134. Chevallier, S., et al., *Application of PLS-DA in multivariate image analysis.* Journal of Chemometrics: A Journal of the Chemometrics Society, 2006. **20**(5): p. 221-229.
135. Zhang, Q., et al., *Serum metabolites as potential biomarkers for diagnosis of knee osteoarthritis.* Disease markers, 2015. **2015**.
136. Villa, J.E., et al., *Fast discrimination of bacteria using a filter paper–based SERS platform and PLS-DA with uncertainty estimation.* Analytical and bioanalytical chemistry, 2019. **411**(3): p. 705-713.
137. Chong, I.-G. and C.-H. Jun, *Performance of some variable selection methods when multicollinearity is present.* Chemometrics and intelligent laboratory systems, 2005. **78**(1-2): p. 103-112.
138. Herberth, M., et al., *Identification of a molecular profile associated with immune status in first-onset schizophrenia patients.* Clinical schizophrenia & related psychoses, 2014. **7**(4): p. 207-215.
139. Miller, B.J., et al., *Meta-analysis of cytokine alterations in schizophrenia: clinical status and antipsychotic effects.* Biological psychiatry, 2011. **70**(7): p. 663-671.
140. Hoirisch-Clapauch, S., et al., *Dysfunction in the coagulation system and schizophrenia.* Translational psychiatry, 2016. **6**(1): p. e704-e704.
141. Cotter, D., et al., *126. Altered Complement and Coagulation Cascade in Age 11 Children Precedes to the Onset of Psychotic Disorder.* Schizophrenia Bulletin, 2017. **43**(suppl_1): p. S69-S70.
142. Rao, W., et al., *A Case-Control Study of the Association between Polymorphisms in the Fibrinogen Alpha Chain Gene and Schizophrenia.* Dis Markers, 2017. **2017**: p. 3104180.
143. Xu, X., et al., *Tubulin β-III modulates seizure activity in epilepsy.* The Journal of Pathology, 2017. **242**(3): p. 297-308.
144. Wang, Y.-C., S.E. Peterson, and J.F. Loring, *Protein post-translational modifications and regulation of pluripotency in human stem cells.* Cell research, 2014. **24**(2): p. 143-160.
145. Smith, L.M. and N.L. Kelleher, *Proteoform: a single term describing protein complexity.* Nature methods, 2013. **10**(3): p. 186-187.
146. Hampel, H., et al., *Total and phosphorylated tau protein as biological markers of Alzheimer's disease.* Experimental gerontology, 2010. **45**(1): p. 30-40.
147. Das, R., A.A. Balmik, and S. Chinnathambi, *Melatonin reduces GSK3β expression and tau phosphorylation via Nrf2 nuclear translocation.* bioRxiv, 2019: p. 861229.
148. Polpitiya, A.D., et al., *DAnTE: a statistical tool for quantitative analysis of-omics data.* Bioinformatics, 2008. **24**(13): p. 1556-1558.

Supplementary data

Supplementary data

Supplementary Table 1. Proteomics data of PBMCs analysis.
ST Dev: standard deviation; CV: coefficient of variation; µg: micrograms; Vol.: volume; [] mg/mL: milligrams per milliliter.

sample	absorvances		mean	St Dev	% CV	µg	Vol. applied	[] mg/mL	50ug (SWATH)	10ug (IDA)
SCZ_4	0.496	0.496	0.496	0.000	0.000	8.882	5.000	1.7763	28.148	5.630
SCZ_5	0.458	0.473	0.466	0.011	2.279	15.254	5.000	3.0508	16.389	3.278
SCZ_6	0.426	0.431	0.429	0.004	0.825	22.984	5.000	4.5969	10.877	2.175
SCZ_8	0.418	0.424	0.421	0.004	1.008	24.551	5.000	4.9103	10.183	2.037
SCZ_9	0.474	0.480	0.477	0.004	0.889	12.851	5.000	2.5703	19.453	3.891
SCZ_12	0.415	0.409	0.412	0.004	1.030	26.432	5.000	5.2863	9.458	1.892
CT_44	0.393	0.392	0.393	0.001	0.180	30.506	5.000	6.1012	8.195	1.639
CT_45	0.504	0.500	0.502	0.003	0.563	7.628	5.000	1.5256	32.774	6.555
CT_48	0.468	0.465	0.467	0.002	0.455	15.045	5.000	3.0090	16.617	3.323
CT_49	0.512	0.512	0.512	0.000	0.000	5.539	5.000	1.1078	45.136	9.027
CT_50	0.433	0.432	0.433	0.001	0.163	22.149	5.000	4.4297	11.287	2.257
CT_51	0.491	0.488	0.490	0.002	0.433	10.240	5.000	2.0479	24.415	4.883
ER	0.417	0.418	0.418	0.001	0.169	25.283	5.000	5.0565	9.888	1.978

Supplementary Table 2. Data analysis of of all 37 proteins that were considered in peptide centered analysis based on 57 relevant peptides.
FC: fold change. VIP: variable importance projection.

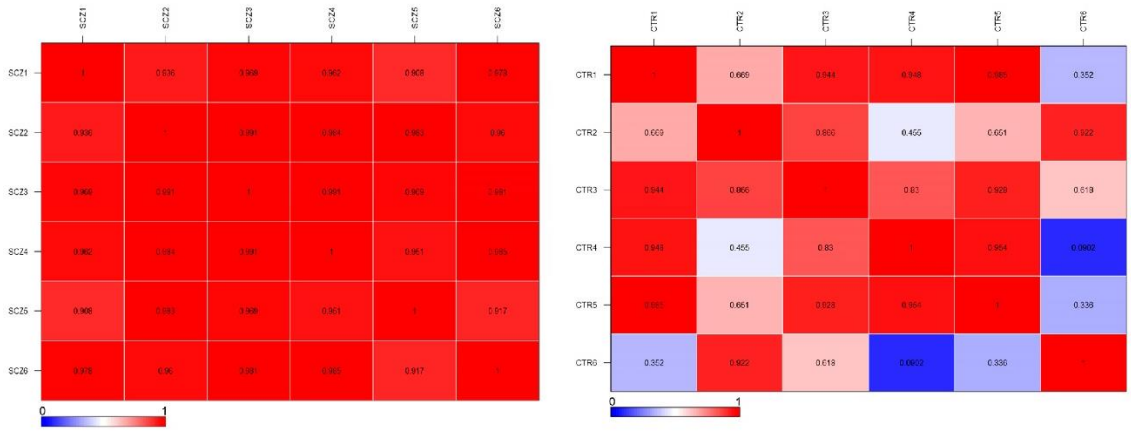
Protein	Peptide	Fold Change peptides	p-value FC peptides	VIP1	VIP2	Fold Change proteins	p-value FC protein	(Pept/Prot) FC	IF (Pept/Prot > |50|% and p-value < 0.05)
sp|O00151|PDLI1_HUMAN	VTPPEGYEVVTVFPK	0.678	0.0281	1.5827	1.6066	0.3721	0.0043	82.18	✓
sp|O00151|PDLI1_HUMAN	SAMPFTASPASSTTAR	0.100	0.0043	1.9868	1.8845	0.3721	0.0043	-73.17	✓
sp|O00160|MYO1F_HUMAN	YFEIQFSR	0.110	0.0087	1.8021	1.7096	0.6861	0.2468	-84.03	✓
sp|O15371|EIF3D_HUMAN	NMLQFNLQILPK	0.259	0.0173	1.8179	1.7653	0.7832	0.2468	-66.93	✓
sp|P02671|FIBA_HUMAN	NNKDSHSLTTNIMEILR	2.496	0.0043	1.8168	1.7225	1.1818	0.7922	111.23	✓
sp|P02671|FIBA_HUMAN	GSESGIFTNTK	0.218	0.0043	2.0376	1.9350	1.1818	0.7922	-81.57	✓
sp|P02671|FIBA_HUMAN	PGSTGTWNPGSSER	0.556	0.0043	1.6958	1.6101	1.1818	0.7922	-52.97	✓
sp|P02671|FIBA_HUMAN	QFTSSTSYNR	0.218	0.0043	2.0188	1.9137	1.1818	0.7922	-81.59	✓
sp|P02675|FIBB_HUMAN	VNDNEEGFFSAR	0.433	0.0087	1.7199	1.6308	1.1447	0.6623	-62.13	✓
sp|P04179|SODM_HUMAN	LTAASVGVQGSGWGWLGFNK	1.656	0.0043	1.8499	1.7830	1.0444	0.4286	58.60	✓
sp|P04406|G3P_HUMAN	VIHDNFGIVEGLM[Oxi]T[Pho]T[Dhy]VHAITATQK	2.834	0.0043	1.7816	1.6894	0.7788	0.1255	263.85	✓
sp|P05106|ITB3_HUMAN	DAPEGGFDAIMQATVCDEK	2.053	0.0222	1.8363	1.7557	1.1851	0.3290	73.26	✓
sp|P07437|TBB5_HUMAN	MAVTFIGNSTAIQELFKR	1.664	0.0173	1.8872	1.8408	0.7449	0.0519	123.40	✓
sp|P08514|ITA2B_HUMAN	GEAQVWTQLLR	1.868	0.0303	1.5222	1.4535	1.1635	0.5368	60.52	✓
sp|P08567|PLEK_HUMAN	SIRLPETIDLGALYLSMK	2.513	0.0043	2.0799	1.9759	1.5975	0.0303	57.29	✓
sp|P08567|PLEK_HUMAN	SIRLPETIDLGALYLSM[Oxi]K	2.989	0.0043	2.0354	1.9411	1.5975	0.0303	87.11	✓
sp|P13639|EF2_HUMAN	GLKEGIPALDNFLDKL	1.688	0.0080	1.7769	1.7539	0.7898	0.5368	113.74	✓
sp|P14618|KPYM_HUMAN	KGVNLPGAAVDLPAVSEKDIQDLK	1.782	0.0087	1.7652	1.6781	1.1870	0.0519	50.12	✓
sp|P14618|KPYM_HUMAN	FGVEQDVDM[Oxi]VFASFIR	2.022	0.0043	1.8781	1.7820	1.1870	0.0519	70.32	✓
sp|P23528|COF1_HUMAN	EILVGDVGQTVDDPYATFVK	0.015	0.0173	1.6718	1.5945	1.2000	0.3290	-98.72	✓
sp|P35579|MYH9_HUMAN	MQQ[Dea]N[Oxi]IQE[KXX]LEEQLEEEESAR	2.437	0.0043	1.8579	1.7815	1.4302	0.0087	70.42	✓

sp|P35579|MYH9_HUMAN	SM[Oxi]EAE[Cox]MIQ[Dea]LQ[Dea]EELAAAER	2.259	0.0043	2.0973	1.9897	1.4302	0.0087	57.92	✓
sp|P35579|MYH9_HUMAN	HEAM[Oxi]ITDLEER	9.918	0.0043	2.1013	1.9983	1.4302	0.0087	593.51	✓
sp|P37802|TAGL2_HUMAN	YGINTTDIFQTVDLWEGK	1.389	0.0043	1.9407	1.8426	0.8594	0.3290	61.66	✓
sp|P37802|TAGL2_HUMAN	NVIGLQMGTNR	0.155	0.0173	1.8764	1.7802	0.8594	0.3290	-81.91	✓
sp|P37802|TAGL2_HUMAN	GASQAGMTGYGMPR	0.106	0.0173	1.8722	1.7758	0.8594	0.3290	-87.71	✓
sp|P37802|TAGL2_HUMAN	QM[Oxi]EQISQFLQAAER	1.698	0.0043	1.6047	1.5217	0.8594	0.3290	97.60	✓
sp|P37802|TAGL2_HUMAN	NVIGLQM[Oxi]GTNR	0.158	0.0303	1.6241	1.5413	0.8594	0.3290	-81.58	✓
sp|P37802|TAGL2_HUMAN	[PGQ]-QM[Oxi]EQISQFLQAAER	2.129	0.0087	1.7481	1.7060	0.8594	0.3290	147.74	✓
sp|P40197|GPV_HUMAN	LM[Oxi]ISDSHISAVAPGTFSDLIK	2.188	0.0043	1.5086	1.4348	1.4321	0.0303	52.76	✓
sp|P49591|SYSC_HUMAN	KEPVGDDESVPEN[Dea]VLSFDDLTADALANLK	1.469	0.0087	1.6602	1.6613	0.8073	0.4286	81.94	✓
sp|P84077|ARF1_HUMAN	MLAEDELRDAVLLVFANK	2.071	0.0043	2.0390	1.9392	1.2735	0.0043	62.62	✓
sp|Q05682|CALD1_HUMAN	LEQYTSAIEGTK	0.172	0.0173	1.8295	1.7350	0.4325	0.0087	-60.25	✓
sp|Q05682|CALD1_HUMAN	[PGQ]-QKEFDPTITDASLSLPSR	0.763	0.0303	1.5222	1.4522	0.4325	0.0087	76.49	✓
sp|Q09666|AHNK_HUMAN	ISMPDVGLNLK	0.156	0.0173	1.7380	1.6687	0.4603	0.0303	-66.08	✓
sp|Q09666|AHNK_HUMAN	VDINAPDVEVQGK	0.165	0.0173	1.7671	1.7723	0.4603	0.0303	-64.19	✓
sp|Q09666|AHNK_HUMAN	VDIETPNLEGTLTGPR	0.086	0.0043	2.0449	2.0182	0.4603	0.0303	-81.21	✓
sp|Q15149|PLEC_HUMAN	MGIVGPEFK	1.770	0.0135	1.6689	1.5831	0.9468	0.7922	86.94	✓
sp|Q15691|MARE1_HUMAN	[PGQ]-QGQETAVAPSLVAPALNKPK	0.439	0.0173	1.7094	1.6238	0.9417	0.7922	-53.33	✓
sp|Q15833|STXB2_HUMAN	EIHLAFLPYEAQVF	2.604	0.0173	1.7885	1.6955	1.5814	0.0173	64.68	✓
sp|Q15942|ZYX_HUMAN	FSPGAPGGSGSQPNQK	0.170	0.0087	1.9551	1.8552	0.3787	0.0043	-55.22	✓
sp|Q27J81|INF2_HUMAN	SVQANLDQSQR	0.444	0.0303	1.6665	1.5879	1.9417	0.0173	-77.14	✓
sp|Q3ZCW2|LEGL_HUMAN	LDDGHLNNSLSSPVQADVYFPR	0.250	0.0080	1.9529	1.8521	0.9589	0.9307	-73.96	✓
sp|Q71U36|TBA1A_HUMAN	AVFVDLEPTVIDEVR	0.200	0.0043	2.1453	2.1146	0.6789	0.0087	-70.51	✓
sp|Q99733|NP1L4_HUMAN	LTDQVMQNPR	0.242	0.0303	1.6653	1.5787	0.9607	0.9307	-74.76	✓
sp|Q9BR76|COR1B_HUMAN	VTWDSTFCAVNPK	1.910	0.0087	1.7154	1.6264	1.1448	0.7922	66.84	✓
sp|Q9BSJ8|ESYT1_HUMAN	LTPRPTAAELEEVLQVNSLIQTQK	2.071	0.0043	1.8399	1.7488	1.2793	0.1255	61.86	✓
sp|Q9BUL8|PDC10_HUMAN	MAADDVEEYMIERPEPEFQDLNEK	1.936	0.0173	1.6341	1.5491	1.1407	0.4286	69.73	✓
sp|Q9NTK5|OLA1_HUMAN	IPAFLNVVDIAGLVK	10.784	0.0043	1.8584	1.7803	1.5734	0.0173	585.39	✓

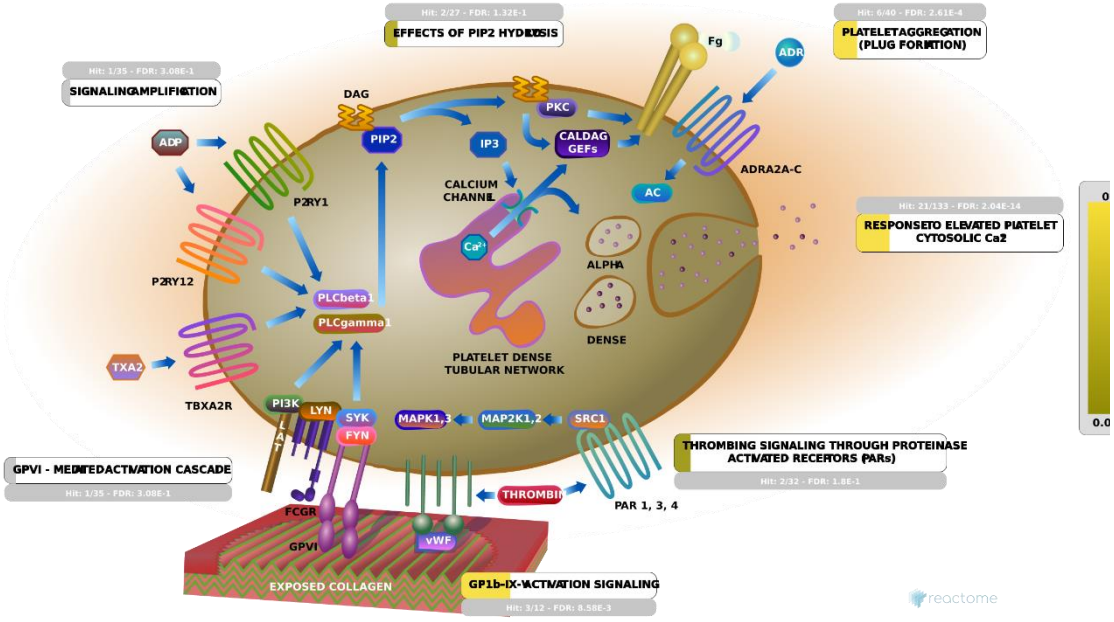
sp|Q9NUQ9|FA49B_HUMAN	DAEGILEDLQSYR	0.032	0.0173	1.6264	1.5623	0.6663	0.1775	-95.25	✓
sp|Q9UBW5|BIN2_HUMAN	TSLEVSPNPEPPEKPVR	0.176	0.0173	1.8078	1.7147	0.8954	0.0519	-80.39	✓
sp|Q9UBW5|BIN2_HUMAN	AGGAAGLFAK	0.099	0.0080	1.9780	1.8800	0.8954	0.0519	-88.92	✓
sp|Q9UBW5|BIN2_HUMAN	ASLGTGTASPR	0.099	0.0087	1.9039	1.8050	0.8954	0.0519	-88.99	✓
sp|Q9Y490|TLN1_HUMAN	SGASGPENFQVGSMPPAQQQITSGQMHR	0.014	0.0087	1.9852	1.8823	1.1833	0.1775	-98.83	✓
sp|Q9Y490|TLN1_HUMAN	FGQDFSTFLEAGVEMAGQAPSQEDR	1.861	0.0043	1.9151	1.8181	1.1833	0.1775	57.27	✓
sp|Q9Y490|TLN1_HUMAN	AQYFEPLTLAAVGAASK	2.077	0.0173	1.5779	1.5004	1.1833	0.1775	75.54	✓
sp|Q9Y5K5|UCHL5_HUMAN	FNLMAIVSDR	0.369	0.0303	1.5047	1.5684	0.7866	0.1775	-53.13	✓

Supplementary Table 3. Pathways found through Reactome Pathway Database.

Pathway name	Entities found	Entities Total	Entities ratio	Entities p-value	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
Platelet degranulation	20	128	0.011	1.11E-16	2.04E-14	6	11	0.001	Homo sapiens
Response to elevated platelet cytosolic Ca2+	21	133	0.012	1.11E-16	2.04E-14	7	14	0.001	Homo sapiens
Platelet activation, signaling and aggregation	29	265	0.023	1.11E-16	2.04E-14	48	115	0.009	Homo sapiens
Hemostasis	37	726	0.063	3.33E-16	4.60E-14	86	333	0.026	Homo sapiens
Smooth Muscle Contraction	8	39	0.003	1.09E-08	1.20E-06	6	11	0.001	Homo sapiens
Vesicle-mediated transport	24	761	0.067	1.14E-06	1.05E-04	78	252	0.02	Homo sapiens
Membrane Trafficking	21	635	0.056	2.85E-06	2.25E-04	70	219	0.017	Homo sapiens
RHO GTPases activate PAKs	5	21	0.002	3.34E-06	2.30E-04	6	15	0.001	Homo sapiens
Formation of Fibrin Clot (Clotting Cascade)	6	39	0.003	4.12E-06	2.51E-04	17	61	0.005	Homo sapiens
Platelet Aggregation (Plug Formation)	6	40	0.003	4.75E-06	2.61E-04	21	27	0.002	Homo sapiens
RAB geranylgeranylation	7	65	0.006	6.50E-06	3.25E-04	2	5	0	Homo sapiens
Integrin signaling	5	28	0.002	1.33E-05	5.57E-04	20	24	0.002	Homo sapiens
Signal regulatory protein family interactions	4	16	0.001	2.76E-05	1.08E-03	3	10	0.001	Homo sapiens
Common Pathway of Fibrin Clot Formation	4	22	0.002	9.41E-05	3.39E-03	10	29	0.002	Homo sapiens
Intrinsic Pathway of Fibrin Clot Formation	4	23	0.002	1.12E-04	3.79E-03	7	24	0.002	Homo sapiens
EPHA-mediated growth cone collapse	4	29	0.003	2.69E-04	8.58E-03	4	4	0	Homo sapiens
RHO GTPase Effectors	11	295	0.026	2.95E-04	8.58E-03	20	113	0.009	Homo sapiens
GP1b-IX-V activation signalling	3	12	0.001	2.96E-04	8.58E-03	6	7	0.001	Homo sapiens



Supplementary Figure 1. Correlation plots between replicates of the same experimental group comparing all quantified proteins performed in InfernoRDN software [148]. Sample CTR6 was excluded from the rest of the analysis since it showed a completely different profile. SCZ= Schizophrenia; CTR= control.



Supplementary Figure 2. Visualization of Reactoms platelet pathways. Overrepresented pathways were platelet activation, signaling and aggregation. A color code indicates the associated pathway p-value.

Annexes

Systematic Review and Meta-analysis

Systematic review and meta-analysis of mass spectrometry proteomics applied to human peripheral fluids to assess potential biomarkers of Schizophrenia

Ana Martinho*, João Rodrigues*, Cátia Santa, Nuno Madeira, Manuel Coroa, Maria João Martins, Antonio Macedo§, Bruno Manadas§

* - equal contribution

§ - equal senior contribution

Background: Neuropsychiatric disorders still demand the discovery of reliable biomarkers capable of predicting and diagnosing these types of diseases.

Mass spectrometry proteomics strategies applied to human peripheral fluids can be a powerful technique to identify those biomarkers.

Aims: To analyze and define the proteomics results common to several studies based on the use of MS applied to human peripheral fluids in individuals with Schizophrenia.

Methods: A systematic review was conducted to compile the reports of proteomics studies applied to human peripheral fluids making use of mass spectrometry strategies in the last 10 years. The outcomes that had enough information available were selected for a meta-analysis.

Results: Our literature search found 19 articles that met the defined eligibility criteria where comparisons between Schizophrenia and controls are studied. Of these, five studies were eligible for a meta-analysis.

Conclusions: Of the five studies that met the criteria to be included in a meta-analysis, six proteins were analyzed. Our results shows that Apolipoprotein C3 is decresased while ficolin-3 is increased in SCZ patients.

1.Introduction

Neuropsychiatric disorders

Psychiatric disorders (PD) comprise a wide range of mental health problems that can severely impact the well-being of those affected [1, 2]. This set of clinical conditions can affect people of different ages and be a main cause of morbidity even in childhood and adolescence [3, 4]. The effects of these disorders on public health are profoundly negative and hugely contribute to the world burden of the disease [1, 3]. About 10% of the world population is affected, with mental disorders making up 30% of the global burden of non-fatal disease (WHO 2016) overcoming cancer and cardiovascular disease and about 1 million people worldwide die annually from suicide [5]. Following the predictions of the World Health Organization (WHO) for the year 2020, depression will be the second leading cause of disability worldwide

Thus, the global situation is bleak with more than 450 million people worldwide living with some form of mental illness, and only in the EU, the number of individuals affected per year is around 165 million [6, 7]. Moreover, it is estimated that one-quarter of the world's population will be manifesting at least one mental disorder at some period of their life [7, 8]. Unfortunately, for several reasons, the progress in understanding PD has been slow [1, 9]

Schizophrenia

The genetic architecture of schizophrenia is extremely complex and heterogeneous and is characterized by rare mutations that are very recent with a relatively high risk and common variants with individually small effects on schizophrenia [10]. Genes implicated by both common and rare alleles operate in pathways that are crucial in brain development, including histone modification, neuronal migration, transcriptional regulation, immune function, and synaptic plasticity [11].

People living with this disease have an average life expectancy significantly reduced, ~20 years lower than the general population. Nonetheless, the mortality rates are high across all age groups [8, 12]. The current diagnosis of schizophrenia is mainly based on phenomenological observation and clinical descriptions using the standard operational criteria defined in systematic classifications, namely the Diagnostic and Statistical Manual of Mental Disorders, edition five (DSM-5), and International Classification of Diseases, version 11 (ICD-11), published by the American Psychiatric Association and WHO,

respectively [3, 13, 14]. The main problem is that these diagnostic definitions have a relatively good reliability, but no established validity.

Epidemiologic studies show that it can take up to several years between symptom onset and diagnosis and the evidence suggests that the earlier the diagnosis, the better the prognosis, by decreasing the duration of untreated psychosis.

The symptoms, which typically arise during adolescence or early adulthood are defined as: (i) positive, such as hallucinations, delusions, and thought disorder; (ii) negative as poverty of speech or alogia, lack of motivation and social withdrawal; and (iii) cognitive symptoms as attention and learning deficits. While positive symptoms can stabilize over the course of the illness, negative symptoms tend to increase and become chronic along with cognitive impairments [15-17].

Psychotic symptoms, which integrate positive symptoms, are a defining feature of SCZ spectrum disorders, and their onset defines the first episode of psychosis [18, 19]. Despite being considered the main feature for the onset and the diagnosis of SCZ, psychotic disorders are characterized by an earlier stage, a pre-psychotic stage termed as prodrome, which is usually missed by clinicians [20, 21].

The treatment of patients is usually based on antipsychotic (AP) medication. After the first successfully employed drug in 1952, chlorpromazine, in the treatment of positive symptoms of SCZ, more drugs were introduced and upgraded in the following years [22, 23]. However, they are still ineffective to around 40% of the patients, and some of them end up discontinuing the treatment or having severe side effects [3, 24]. The rates of comorbid illnesses associated with SCZ are high, with patients usually linked to an increased metabolic syndrome (MetS) risk, fixed on 32.5% in SCZ patients in a study by Mitchell et al. 2013 [12].

The pathophysiology of SCZ remains unclear, lacking a comprehensive view of the underlying neurobiological mechanisms, although some aspects are beginning to be clarified. Dopaminergic dysfunction has been one of the pathophysiological hypotheses defended for decades, under various formulations and which is also supported by the most recent genetic findings.

Hypo and hyperactivities of the dopaminergic system are seen in SCZ patients, and both are linked to the symptoms previously described [25, 26]. Additionally, other dysfunctions underlying the pathophysiology of SCZ, such as neurotransmitter signaling of glutamate, hypothalamic-pituitary-axonal (HPA) axis signaling, and immune system dysregulation, as well as synaptic plasticity anomalies, have been reported [16, 26, 27]. Changes in brain

structures, which have also been proposed as etiologically relevant, are correlated with some of these alterations [27].

Despite the efforts to elucidate the mechanisms or etiology behind neuropsychiatric disorders, they remain elusive and not yet clarified. As biomarkers can reflect changes upon CNS diseases, namely the dysregulation of molecular expression profiles, the need to search for reliable biomarkers is becoming imperative, hopefully improving the misdiagnosis of patients [3].

The search for biomarkers

To improve the knowledge about these complex disorders, "omics" approaches have emerged to shed light on disease pathogenesis and support a trustworthy way of prediction and diagnosis for PD [17, 28]. With a huge potential associated, high-throughput omics technology can be a solution to predict clinical endpoints, being the improvement of patient care and outcome the ultimate goal. However, the translation from research to a successful clinical omics-based test is far from the great potential of these approaches [29, 30].

The search for candidate biomarkers is one of the outputs of "omic" studies. According to the National Institute of Health (NIH), a biological marker, generally just termed as a biomarker, is a "characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [31]. The study of the brain and the associated disorders is complex since it presents a high degree of inter- and intra-cellular heterogeneity and so, different locations may have a distinct proteome as a consequence of alterations in different cell types and cellular networks. The proteome of the Central Nervous System (CNS) can change even with minimal alterations in the normal course of its development and/or function [32, 33]. To understand the alterations and the mechanisms related to a disorder, we should analyze qualitative and quantitative changes in the complete set of proteins encoded by an organism genome at different or specific points in time [18, 34]. Proteomics can be a powerful tool since it can give a real-time evaluation of an individual state, health vs. disease, and, in an ideal scenario, predict the susceptibility to develop a specific mental disorder [4, 32]. The possibility of not only identifying but also quantifying the proteins makes the proteomic approach more reliable to evaluate psychiatric diseases at different levels. Moreover, protein-based tests can offer the nearest view of the pathophysiological process behind the PD since their expression and function are the result of what happens during post-transcriptional (e.g., alternative mRNA splicing) and post-translational events

(e.g., phosphorylation, glycosylation, oxidation), as well as the interactions between them [3, 4, 35].

The discovery of biomarkers in neuroscience is challenging but may help to reveal disease-related alterations and consequently improve clinical settings, for instance, helping to predict diagnosis, even before the onset, patient stratification, and monitoring of disease progression and treatment [28]. Early and guided interventions will improve patients' outcome as they are usually prescribed with a medication that will not elicit a proper response or will be ineffective, and it will have to be altered until a desired response is achieved (trial-and-error testing). Moreover, a change in the disease diagnosis is also common. Therefore, it would increase the quality of life of individuals and reduce the burden associated with psychiatric disorders, namely from misdiagnosis, high rates of hospitalization, and treatment expenses, which have a huge impact on health costs [3, 36, 37].

Biological markers in psychiatric disorders

The search for biomarkers in psychiatric disorders began with post-mortem brain tissue and CSF. In contrast to body fluids, brain tissue can only be accessed during autopsies, not being useful for disease diagnosis or to be used in longitudinal studies. Additionally, some common variables and confounding factors, such as post-mortem interval and pH range, can impact this tissue's integrity. The contribution to protein degradation, as well as medication and age, also lead to drawbacks that cannot be avoided [18, 38]. More recently, the whole-body concept emerged since the integration of the brain and a variety of physiological conditions are now known to be reflected in the contents of body fluids [18, 39]. This link created between the brain and the periphery enhanced the search for biomarkers in body fluids that could be easily accessible, like blood [3].

Mass spectrometry

Since its development, mass spectrometry (MS)-based technologies have been improved and, in the last decades, became a well-suited method for biomarker discovery, supporting the expansion of the proteomics field [40, 41]. The success of MS in proteomics is due to its specificity and sensitivity, which are mainly attributable to advances in liquid chromatography in tandem MS (LC-MS/MS) approaches. This type of technology is capable of revealing proteome insights at the level of composition, structure, and function. Proteomics tools make possible the qualitative and quantitative (either relative or absolute) analyzes of proteins in complex biological samples [42, 43].

In the beginning, successes in proteomics approaches were supported by two-dimensional electrophoresis (2-DE), with complex protein mixtures being separated by its molecular charge (isoelectric point) and mass (molecular weight) in the first and second dimension, respectively. In this approach, protein abundances were calculated based on stained protein spots' intensities, followed by MS analysis for protein identification [18, 44, 45]. Although improvements were made, other methodologies emerged to circumvent some of the previous technical drawbacks, namely to face the dynamic range limitations and the unsuitable separation and detection of some protein subtypes, as membrane proteins [18, 42]. Throughout the years, improvements in proteomics approaches were made, and a variety of more in-depth MS-based methods were quickly applied to compare protein profiles, usually between control versus disease states. Considering this, there are two main groups within quantitative proteomics methods: i) labeling techniques, which involve different isotopic labeling of samples, including chemical, enzymatic or metabolic labeling, followed by MS analysis; and ii) label-free techniques, where the sample is individually analyzed without the addition of any other chemical compound. The last quantitative approaches are regarded as a versatile and cost-effective alternative approach to labeled quantitation and have gained significant interest in recent years, mainly due to the development of more sensitive and reliable methods. Additionally, some methods capable of detecting either relative or absolute peptide levels can also provide a targeted MS approach and be used as a validation method [45-47].

This article provides a systematic review and meta-analysis on the use of MS-based methods in proteomic studies to assess biomarkers or a panel of biomarkers associated with Schizophrenia (SCZ) based only on the analysis of peripheral fluids in the last 10 years.

METHOD

As this study used systematic review and meta-analysis strategies, ethical approval for this study and informed consent statement is not required. We included all articles that met all the keywords that specified the objective of the study.

Search strategy

Articles included in the systematic review were identified through a computer-based search conducted from January 2010 to December 2020 in two independent databases: PUBMED and Web of Science. The search was conducted using the following keywords: SCHIZOPHRENIA AND PROTEOMIC* AND MASS SPECTROMETRY, (**Figure 1**). Searches

were restricted by language (English). Two distinct observers performed the literature search independently to identify articles that potentially met the inclusion criteria, and disagreements were discussed between authors and were resolved by group discussion. Extracted data were entered into a computerized spreadsheet for analyses. Then, the reference lists of the included studies were scrutinized, excluded studies, and previous reviews were searched. The study authors were contacted to request additional information when necessary.

Eligibility criteria

Studies were included if they met the following criteria: (a) the manuscripts had to be published in English, and only peer-reviewed published research papers could be incorporated; (b) the research design included the use of mass spectrometry techniques for proteome profiling and/or quantification; (c) studies performed in human peripheral fluids samples; (d) the research design included a control group comprising healthy volunteers or the study of the effects of medication in the same cohort of SCZ patients before and after treatment; (e) publication between 2010-2020, (f) for the meta-analysis the study statistics were convertible to effect size (e.g., means and standard deviation, F- or t-values or exact p-value).

Data extraction

Two authors independently extracted the following data from the eligible studies, according to a pre-specified protocol of data extraction (**Table I and Table II**): (1) authors; (2) DOI; (3) year of publication; (4) participants characteristics (including diagnosis type, sample size and group comparison, mean age, mean illness duration, gender, medication status, type of peripheral samples, and clinical criteria applied); (5) analytical technique; (6) sample preparation (protein depletion or/and enrichment); (7) differences between protein levels of SCZ patients as measured against controls or other mental disorders; and (8) altered pathways.

Any discrepancies between the extracted data were resolved in a group meeting.

Statistical Analysis

RESULTS

Characteristics of included articles

The selection of eligible studies included in our systematic review is shown in **Figure 1**. From the searches performed in WOS and PUBMED databases, a total of 286 potentially

relevant research manuscripts were identified in the initial screening. Additional studies are not included since articles that were identified as similar in both databases and accessed manually did not fit our research aim. Based on abstract review and exclusion of duplicates, 133 were retrieved for more detailed evaluations. Of these research manuscripts, 36 were excluded after full-text reading (the abstract reading was not enough to exclude these articles immediately, and despite the match of keywords, the studies did not fit on the inclusion criteria), and 78 were identified as reviews or studies that used mice/rats, cell lines, and brain samples. In total, 19 papers met all eligibility criteria and were included in this systematic review. From these studies, some reached the criteria to be included in the meta-analysis.

The study characteristics for each of the 19 studies are shown in Table I. As mentioned, only studies of human peripheral fluids were considered.

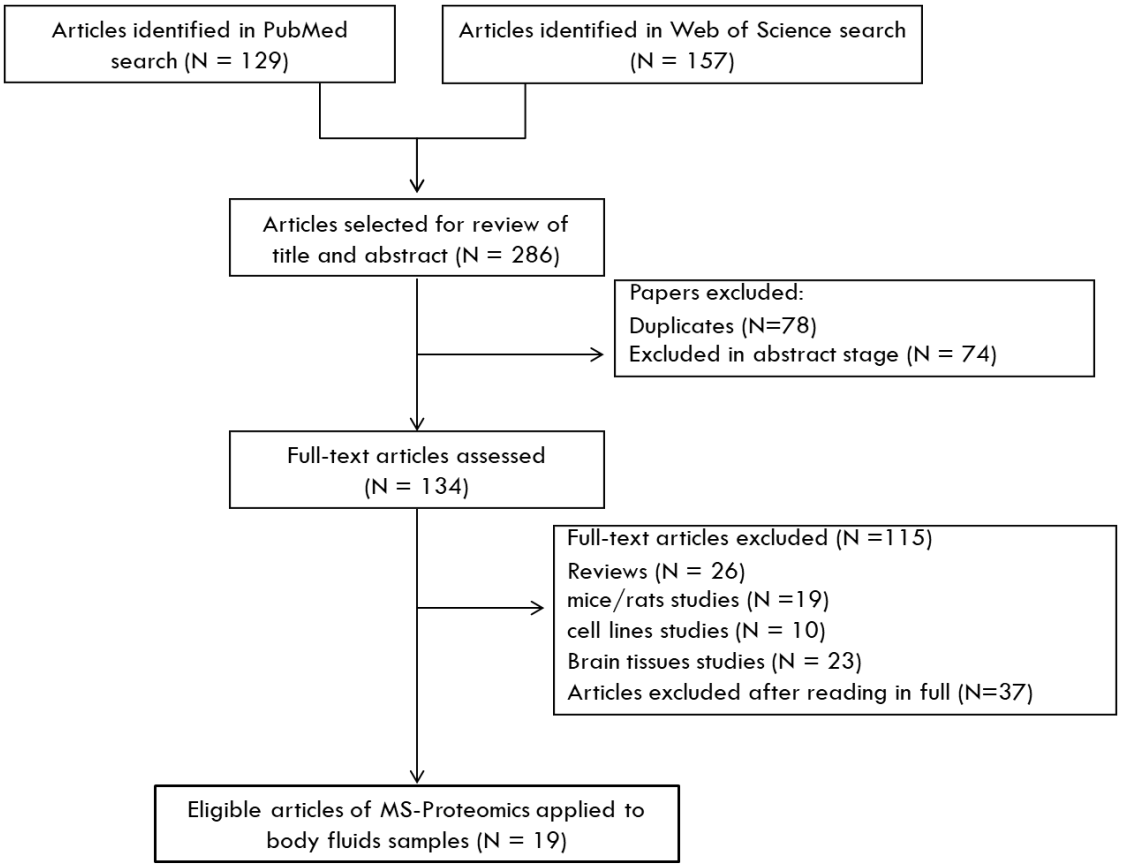


Figure 1 Flow chart of the selection process of the studies included in this systematic review of peripheral fluids MS-based Proteomics in SCZ disorder.

All articles included in this review were published from 2010 and onward, being the reflection of the most recent proteomic works in the context of biomarker findings in SCZ and biological pathways changes produced in the last ten years. Additionally, Table II shares

a list of more detailed information for each study as the techniques applied for the samples in the study and the respective findings.

Number of samples

The number of patients with SCZ included in the studies varied between 4 [48] to 80 [49]. When compared with the first years of the decade in analysis, the last years of publications show an increase in the number of patients per study (see **Figure 2**). Two of the studies with the higher cohort of individuals with SCZ were recently published, 2017 [50] and 2019 [51], and both with n=60. In 2017, three out of four published studies have a cohort of SCZ composed of less than 30 individuals [52-54]. On the other hand, of the four studies published in 2019, three of them studied more than thirty individuals with SCZ [51, 55, 56]. Until 2012, all studies used a number of individuals with SCZ lower than thirty (n<30). Since 2014, thirteen studies were published, and more than half studied more than 30 individuals with SCZ. With this information, it is clearly noticeable that more recent studies privilege the use of larger cohorts, which is a good parameter to achieve a significant result.

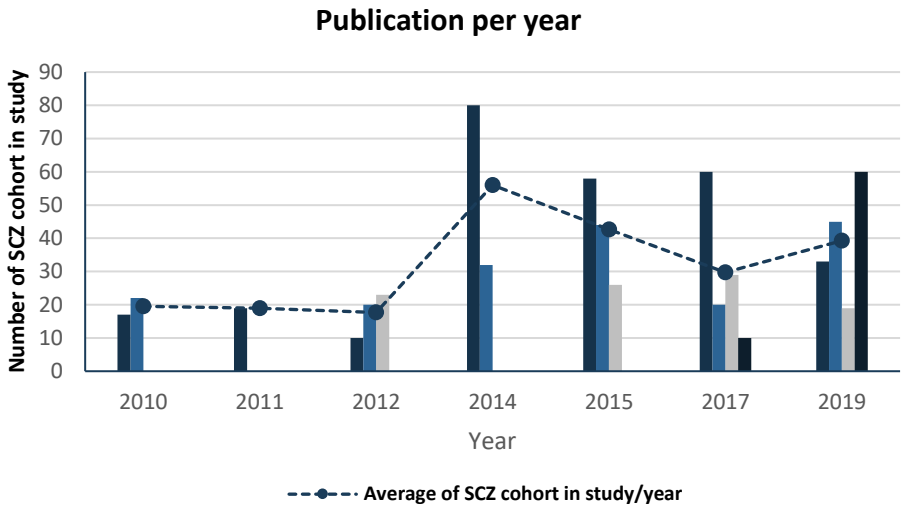


Figure 2- Publication frequency. The number of bars shown in the graphic reflects the number of articles published per year over the last ten years, and the height of each bar reflects the number of SCZ patients in the cohort of the study. The average number of SCZ patients in the cohorts per year are shown in the markers connected by the dashed line.

Diagnostic criteria

It stands out that DSM-IV was the most used diagnostic criteria (11 studies), and two of these studies also used ICD (see Table I) as the diagnostic criteria [57, 58]. ICD-10 was the second most used criteria (6 studies) and applied together with ICD-9 in one study [50]. DSM-V, the last DSM manual available, was only used in one study published in 2019 [56]. There is only one study that does not share this information [48].

Age

Throughout the studies, the reported average age of the studied individuals was comprised between 16.25 [59] and 51.9 [60] years, and only two studies reported an average age below 25 years [49, 59]. The majority of the studies have an average age between 29 and 43 years (15 studies) for the SCZ group; however, one study has no information about the age of the individuals within the study [58]. In concordance with the information provided by the studies that usually try to make the comparison between age-matched groups, the average age was similar in the different groups in the studies.

Gender

Considering the gender information in the SCZ group (see Table I) provided for 18 of the 19 studies in this analysis, only 3 studies (all published in 2012) used a proportional number of samples, with 2, 5, and 10 samples representing each gender [59-61]. Male gender patients are prevalent [48-52, 54, 56, 57, 62-65], sometimes 2-3 times higher than the samples of the female gender in the study. Only three studies had a larger number of samples for the female gender than the male gender [53, 55, 66], and two of them with a minimal difference.

In the control group, the differences of the gender are not so exacerbated but follow the same trend, with samples of the male gender being more represented. Only two studies that used a control group did not provide gender information [58, 59]. Once more, three articles used the same number of samples representing each gender, but only two of them match the number of samples and gender used in the SCZ group [60, 61].

Considering the four studies with a group of other disorders (BD and depression), two studies had a higher number of males [52, 55] and two a higher number of females [62, 66]. Only two of them follow the same gender ratio in the SCZ and CTL groups [52, 66].

Illness duration

Only a small number of studies refers to the illness duration (all in plasma and serum). Of the seven studies that contain this information, only one used individuals with SCZ within the first year of diagnosis (the same study intended to study individuals with first episode of psychosis) [49]. In the other studies, the illness duration is comprised between seven and twelve years [48, 52, 55, 56, 62, 65]. Considering all studies published in 2019, only one of them did not report this information [51], which contrasts with none of the studies published before 2014 containing this data. Although it is more common to see this

information reported in the studies published in the last 5 years, there is a clear need to standardize the type of data reported in these studies.

Table I – Demographic information of all studies included in the systematic review of Schizophrenia and biomarkers discovery using MS-based method in human peripheral fluids.

First authors	Year	Cohort	Schizophrenia (SCZ)				Controls		Other diseases			Ref.
			Clinical criteria	Age (years)	Illness duration (years)	Gender M/F	Age (years)	Gender M/F	Cohort	Age (years)	Gender	
Levin Y	2010	22 SCZ 33 CTL	DSM-IV	29.0	-	15;7	28	18;15	-	-	-	[63]
Martins-De-Souza D	2010	17 SCZ 10 CTL	DSM-IV and ICD10	31.2	-	11;6	30.8	5;5	-	-	-	[57]
Herberth M	2011	19 SCZ 19 CTL	DSM-IV	29.7	-	14;5	34.5	12;7	-	-	-	[64]
Raiszadeh MM	2012	23 SCZ; 55 SCZ (only 4 of each group were used)	DSM-IV	16.25	-	2;2	±22	2;2	-	-	-	[59]
Jaros JA	2012	20 SCZ 20 CTL	ICD-10	31.4	-	10;10	32.1	10;10	-	-	-	[61]
Li Y	2012	10 SCZ; 10 CTL	DSM-IV	51.9	-	5;5	52.6	5;5	-	-	-	[60]
Iavarone F	2014	32 SCZ 31 CTL (12 smokers	DSM-IV and ICD	-	-	-	-	-	17 BD	-	-	[58]

		and 19 non-smokers)										
Song X	2014	80 SCZ	DSM-IV	24	11 (months)	42;38	-	-	-	-	-	[49]
Al Awam	2015	26 SCZ 26 CTL	DSM-IV	31.7	12.12	20;6	37.04	20;6	-	-	-	[65]
Ding YH	2015	44 SCZ 40 CTL	ICD-10	33.1	-	20;24	34	18;22	26 DP	32.5	11;15	[66]
Martins-de-Souza D	2015	58 SCZ	-	36.81	6.15	35;23	-	-	-	-	-	[48]
Alekseeva	2017	10 SCZ 10 CTL	ICD-10	35	-	6;4	39	3;7	-	-	-	[54]
Knöchel C	2017	29 SCZ 93 CTL	DSM-IV	37.16	11.81	21;8	33.59	44;39	25 BD	37.39	19;6	[52]
Huang TL	2017	20 SCZ 20 CTL	DSM-IV	38.15	-	9;11	38.5	7;13	-	-	-	[53]
Cooper JD	2017	60 SCZ 77 CTL	ICD-9/ICD-10	30.1 (M) 31.8 (F)	-	31;29	31.1 (M) 32.7 (F)	43;34	-	-	-	[50]
Walss-Bass C	2019	60 SCZ 20 CTL	DSM-IV	42.5	-	46;14	41.1	14;6		-	-	[51]
Pessôa GS	2019	19 SCZ 13 CTL	ICD-10	37	7.6	13;6	38	3;10	19 BD	41	7;12	[62]
Rodrigues-Amorim D	2019	45 SCZ 43 CTL	DSM-V	40.78	11.63	28;17	43.82	23;17	-	-	-	[56]
Smirnova L	2019	33 SCZ 24 CTL	ICD-10	34 ‡	7 ‡	11;22	28 ‡	6;18	23 BD	32 ‡	14;9	[55]

SCZ= Schizophrenia; CTL= control; BD= Bipolar Disorder; DP= Depression; M= male; F= female; ‡ - median value

More detailed information of the studies is summarized in Table II with the indication of biological sample and type of sampling, cohort information, diagnostic criteria, treatment information (treated or drug naïve), type of MS-based method, other techniques applied, use of depletion and/or enrichment, differentially expressed regulated proteins identified, and altered pathways and major findings.

Cohort information

Seventeen studies focused on the search of diagnostic biomarkers since they make a comparison between a group of individuals with SCZ and a control group. Within these, there is one interesting study that used a control group that distinguishes smokers and non-smokers healthy individuals [58]. The two other studies included in this review do not have a control group but used; instead, two groups of individuals with Schizophrenia, treated and not treated (drug naïve), and aimed to study the difference between both [48, 49]. Thus, some studies also try to highlight the effects of psychotropic drugs on the therapeutics of SCZ. Additionally, despite the fact that most studies have analyzed a cohort of SCZ against a control group (healthy individuals), some of them also used a group of patients with order disorders, specifically, Bipolar Disorder [52, 55, 58] and Depression [66] (see Table I). For one of them, the cohort of SCZ was compared together with the cohort of BD against the control group [52], while in another study, the SCZ group was compared against the BD and CTL groups [55].

Frequency of the publications

It can be observed that the publication of studies within this subject does not follow a pattern throughout the last ten years (**Figure 2**). More precisely, in 2013, 2016, and 2018, we could not find any articles that fit our research strategy. However, it is noticeable that there is an increase in the number of publications over the years.

Type of sample and sampling

Based on the studies encountered when searching the databases (**Figure 1**), we were able to notice that brain tissue is still being analyzed, but the number of studies in body fluids is increasing (see **Figure 3**), and it provides some evidence that it can be a reliable choice for biomarkers research [67]. Biofluids are suitable matrices that enable more user-friendly tests, as the majority of them are of easy access [68, 69]. Based on its accessibility, they can be categorized as non-invasive (saliva, sweat, urine, and tears), minimally invasive (blood), and invasive (cerebrospinal fluid).

Due to its proximity to the brain, cerebrospinal fluid (CSF) is considered relevant in the study of brain disorders. Despite being a dynamic fluid, CSF has to be collected through a lumbar puncture, which is an invasive procedure and leads to a minimal amount of fluid, limiting the possibility of this type of analysis in CSF [18, 28]. As it can be inferred from this review (see **Table II**), with these search parameters, only one study in CSF was retrieved [57]. This shows the importance of selecting more readily accessible samples [18]. Studies with plasma and serum samples have increased over the last years and looking at psychiatric disorders as whole-body diseases, somehow, contributed to this change [69]. Besides protein content being significantly more abundant than what is found in CSF, approximately 500ml of CSF are exchanged daily to circulating blood [40, 41]. Additionally, dynamic changes can also be studied in these samples, which can be collected in reasonable amounts and by straightforward and safe procedures [28].

Among these nineteen studies, the most prevalent sample type studied was serum (n=8), followed by plasma (n=6) (**Figure 3**). Since 2015, this prevalence was kept with serum and plasma, with only one study using another type of sample, which was PBMCs [53]. Two of the studies used PBMCs [53, 64], and only one study used CSF [57], sweat [59], or saliva samples [58] (2010, 2012, and 2014, respectively). As it can be observed (**Table II** and **Figure 3**), plasma and serum are the main samples used among the studies. These biofluids can be easily sampled and have been widely used in proteomics-based research and for disease diagnosis. However, the complexity and dynamic range that characterize the proteome of both samples are responsible for unsatisfactory outcomes in the search for disease biomarkers. As potential biomarkers are usually present in low concentrations and tend to be masked by high abundant proteins, some strategies have been applied to overcome this challenge. Depletion of high-abundance proteins and enrichment of low and medium abundance proteins are two methods used to circumvent this problem [70]. As we can observe in Table II, most of the studies where the sample in use was plasma or serum performed the depletion of high-abundance proteins. In fact, the studies that analyzed plasma (n=6) did not show a prevalence about making the depletion or not, while most of the studies based on serum samples (five in a total of eight) applied the depletion of high-abundant proteins. In a total of six studies that used enrichment techniques (proteominer, aptamers, IMAC, IMAC30, C18 TiOtips, subcellular fractionation) only one was not analyzing serum or plasma but PBMCs, instead [64]. Among the 14 studies that analyzed plasma and serum, only 3 studies did not use depletion or enrichment techniques [50, 52, 62].

Only four studies worked with a pool of samples. Two of them were published in 2019 and analyzed plasma and serum [51, 62]. The other two analyzed CSF and sweat samples and were published in 2010 and 2012, respectively [57, 59].

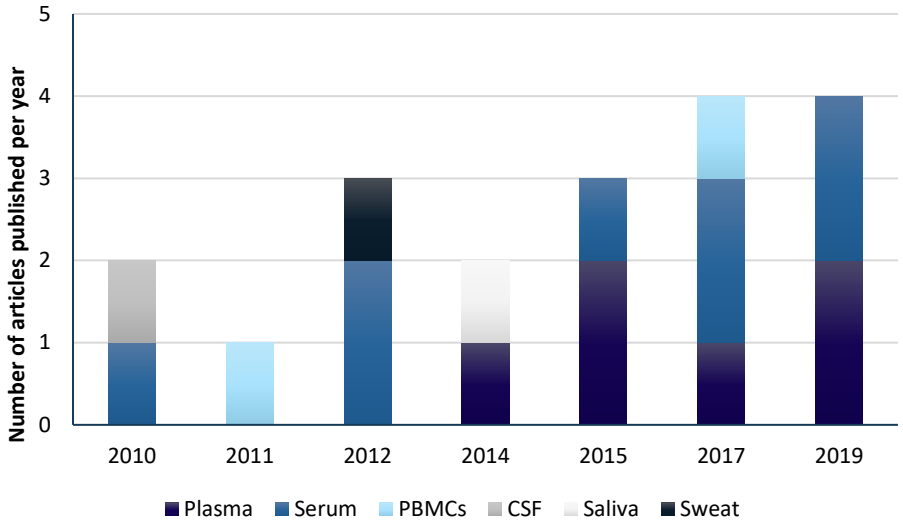


Figure 3– Sample type. The image shows the number of publications per year that fit the criteria of this review. Each color shows the type of samples used, and its height indicates the number of studies published.

Drug naive

Regarding being treated or not (drug naïve/minimally medicated), we noticed that there is not a prevalence of one over another among the studies and only one does not share this information. Additionally, two studies worked exclusively with a single cohort of individuals with SCZ, which were analyzed before and after treatment [48, 49].

MS-based methods

Between 2010-2015, there is a prevalence of the use of LC-MS/MS methods in the studies (n=7), while the other studies applied MALDI TOF/TOF analysis [49, 57, 65, 66]. In the next years, an increase in studies that applied in-gel digestion is observed, as we can see from the articles published in 2019 (n=4), followed by LC-MS/MS analysis [51, 55, 56]. It is undeniable that LC-MS/MS analysis was the prevalent MS-based technique among the 19 studies considered in this review.

Additionally, one study applied ICP-MS-based methods to assess the interactions between metals and proteins [62].

Validation

Immunoassay methods have been applied in the validation of specific proteins identified as differentially expressed. More than half of the studies used immunoassays methods for validation of protein's expression pattern, being enzyme-linked immunosorbent assay

(ELISA) the most used approach [49, 51, 55, 60, 61, 63], followed by Western blot (WB) [48, 56, 57, 64] and MRM (Multiple Reaction Monitoring) [59].

Before 2015, the immunoassay of choice for validation was ELISA. In the following years, neither ELISA nor WB were predominant over the other. Also, among the studies published in 2019, only one does not apply a validation technique [62].

Table II. Proteomic studies of schizophrenia biomarker discovery using MS-based method in human peripheral fluids. The proteins identified as altered are represented by their entry name as described in UniProt (the corresponding accession name and protein name are described in **Table S1** and **Table S2**).

Year	Cohort information	Samples	Sampling	Treatment	MS-based method	Other techniques	Depletion/ enrichment	Quantification method	No.proteins	Altered proteins	Altered pathways	Ref.
2010	22 SCZ 33 CTL	Serum	Individual	Treated	LC-MS/MS	ELISA	Yes/No	MS	Total ID: 1411; significantly different: 10	CD5L ↓ ; IGHM ↓ ; F13B ↓ ; TRFE ↓ ; APOD ↓ ; APOA1 ↓ ; FETUA ↓ ; APOA4 ↓ ; APOA2 ↓ ; APOC1 ↓	Lipid metabolism; molecular transport; immune response	[63]
2010	17 SCZ 10 CTL	CSF	Pooled	Drug naive	2DE MALDI-TOF/TOF MS	WB	No/No	2DE	Total ID: 6	APOE ↑ ; PTGDS ↑ ; APOA1 ↑ ; TTHY ↓ ; TGFR1 ↓ ; CCDC3 ↓ ;	Lipid metabolism	[57]
2011	19 SCZ 19 CTL	PBMCs	Individual	Drug naïve/treated	LC-MS/MS	WB	No/Yes	MS	Total ID unstimulated: 185; Stimulated: 441; Significantly different: 19 between drug naive SCZ and CTR	Unstimulated PBMCs: CNDP2 ↑ ; Uncharacterized protein KIAA0423 ↑ ; LDHB ↑ ; COTL1 ↓ ; GPI ↓ ; HSP72 ↓ ; Stimulated PBMCs: ALDOC ↑ ; GAPDH ↑ ; HNRPK ↑ ; LDHB ↑ ; MYH14 ↑ ; MYH15 ↑ ; NAMPT ↑ ; PGK1 ↑ ; PPIA ↑ ; TPIS ↑ ; PKLR ↑ ; PGAMA4 ↑ ; CH60 ↓ ;	Glycolytic pathway, Immune response	[64]
2012	23 SCZ 55 CTL For analysis: 4 SCZ; 4 CTL (2nd pool)	Sweat	Pooled	Treated	LC-MS/MS and LC-MS/MS-MRM		No/No	MS	1st set Total ID: 150; 2nd set Total ID: 185; MRM: 30	ZA2G ↑ ; ANXA5 ↑ ; ARG2 ↑ ; BLMH ↑ ; CALL5 ↑ ; CASPE ↑ ; CDSN ↑ ; CSTA ↑ ; DCD; Desmoglein ↑ ; DJ-1 ↑ ;G3PDH ↑ ; KLK11 ↑ ; KRT10; PRDX1 ↑ ; PEBP1 ↑ ; S100A7 ↑ ; THIO ↑ ; PIP ↓ ;	Metabolic process	[59]

2012	20 SCZ 20 CTL	Serum	Individual	Drug naïve	LC-MS/MS	ELISA	Yes/Yes‡‡	MS	Detected: 694; Total ID: 312; Significantly different: 35; Phospho altered: 72	K2C6B ↑; FCN3 ↑; SRBS1 ↑; NUCB1 ↑; K1C9 ↑; NUDT6 ↑; ALS2 ↑; IBP3 ↑; MAST1 ↑; CFAB ↑; C4BPA ↑; FHR3 ↑; ITIH3 ↑; CO6 ↑; AGRE1 ↑; CAH1 ↓; RET4 ↓; LRRC7 ↓; FR1L6 ↓; KI21B ↓; TETN ↓; KIF27 ↓; APOA1 ↓; APOA2 ↓; MYOF ↓; FIBA ↓; CCD57 ↓; SMC1A ↓; K1C14 ↓; PHLD ↓; LIFR ↓; XIRP1 ↓; WDR19 ↓; SMC4 ↓; SAGE1 ↓	Acute phase; Complement and coagulation system; Immune Response	[61]
2012	10 SCZ 10 CTL	Serum	Individual	Drug naïve	LC-MS/MS	ELISA	Yes/No	MS	Total ID: 1344; 192 used for PLS-DA (27 SCZ related)	CO8B ↑; CD5L ↑; DOPO ↑; IGHG4 ↑; IGHM ↑; KNG1 ↑; PI16 ↑; PGRP2 ↑ ITIH4 ↑; PLTP ↑; IPSP ↑; IGK@ protein ↑; IGL@ protein ↑; AMPN ↓; APOC2 ↓; APOF ↓; C4BPB ↓; APOL1 ↓; FA7 ↓; GGH ↓; ICAM2 ↓; ALS ↓; isoforms 2 of ITIH4 ↓; LBP ↓; PROS ↓; ZNF57 ↓ (SCZ related proteins)	Complement cascade pathway	[60]
2014	32 SCZ 17 BD 31 CTL (smokers and non-smokers)	Saliva	Individual	-	LC-MS/MS	-	No/No	MS	Total ID: 8	DEF1 ↑; DEF2 ↑; DEF3 ↑; DEF4 ↑; S10AC ↑; CYTA ↑; CYTB ↑; CSTB ↑; (Patients – SCZ and BD vs Control)	Innate Immunity	[58]
2014	15/80 SCZ (proteomic analysis of 15 randomly selected patients; all 80 patients were used	Plasma	Individual	Drug naïve vs. treated	2DE MALDI-TOF/TOF	ELISA	Yes/No	2DE	Total ID and altered: 18	APOA1 ↑; C4B ↑; CFB ↑; NEB ↑; C8B ↑; ZN185 ↑; PLMN ↑; HEMO ↑; HNMT ↑; GBP1 ↑; FGG ↑; TRFE ↑; ALEX ↓; RET4 ↓; K1C9 ↓; K2C1 ↓; VINC ↓; GELS ↓	Metabolism	[49]

	in ELISA analysis)											
2015	26 SCZ 26 CTL	Plasma	Individual	Treated	MALDI-TOF-MS		No/Yes	MS	Total Detected: 94; Significantly different: 11 protein ions from TiO; and 5 from C8	m/z 3177 suggested to be a fragment of Apolipoprotein A1	-	[65]
2015	44 SCZ 26 DP 40 CTL	Serum	Individual	Treated	SELDI-TOF-MS and MALDI-TOF MS		No/Yes	MS	Significantly different: 91 protein peaks	N-terminal fragment of fibrinogen ↓	-	[66]

2015	58 SCZ	Plasma	Individual	Drug naïve/free vs treated	LC-MS/MS	WB	Yes/No	MS	Total ID: 98541 differentially expressed proteins in responders (t6-R/t0-R); 58 differentially expressed proteins in non-responders (t6-NR/t0-NR)	Responders: COR2A ↓; CCN1↓; EFCB6 ↑; FARP2↑; PRKDC↑; SNX17↓; TGFI1↓; THS7A↓; CARL2↑; FA9↓; HS71L_HUMAN↓; ITIH3↑; RMO3↓; AACT; U2QL1↓; MLH3↓; MYCB2↑; MYCD↓; P73↓; ZN215↓; ABCB5↑; ABCAD; ALBU↓; APOA4↓; VDAC1↑; COPA1↓; ECM1↑; ELMO1↓; NEBU↓; TBG1↓; CD5L↑; CF157↓; CCD71↑; CC74A↑, CENPP↑; FDX2↑; YOOO3↓; LIX1↓; 5NT3L↑; PDZD2↑; TMCO2↑; Non-responders: CMGA↓; COR2A↑; CCN1↓; EFCB6↓; FARP2↑; MUC16↓; OR4KD↓; PRKX↑; RNF11↑; SEPT7↑; SNX17↑; TEX14↑; TGFI1↓; WEE1↓; THS7A↓; MLH3↓; PARP2↑; DDX46↑; DNMT1↓; MYCD↑; PCBP2↓; ZC3HD↓; ZFHX4↑; ZN215↓; ABCAD↓; ALBU↓; APOA4↓; APOB↑; GOGA4↑; S14L4↑; SYN3↓; VDAC1↑; VOPP1↑; ECM1↓; KIF17↓; KIF3A↓; MVP↑; MYPC1↓; TBG1↓; SAMP↑; CLPX↑; FA9↑; PROS↑; U2QL1↑; CO2↑; UTY↑; TITIN↑; K1109↓; CI117↑; CCD71↓; ECT2L↓; FDX2↑; FIGL2↓; GDPD5↑; 5NT3B↑; SPT2↓; TMCO2↑; WDR11↑;	Cell communication and signaling; Protein metabolism; Regulation of nucleic acid metabolism; Transport; Cell growth and maintenance; Immune response; Energy metabolism	[48]
2017	10 SCZ; 10 CTL	Serum	Individual	Drug naive	2DE MALDI-TOF/TOF		Yes/No	2DE	Total ID and altered: 15 proteins	↑APOA4; ↑HP; ↑↓APOC3; ↓SAA1; ↓CLU; ↓TTR; ↓Albumin fragment (protein ID); ↓alpha2-antitrypsin (protein ID); ↓Haptoglobin hp2α (protein ID);	-	[54]

2017	29 SCZ 25 BD 93 CTL	Plasma	Individual	Treated	LC-MS/MS-MRM	-	No/No	MS	42 plasma proteins were quantified and analyzed	(average)	ApoC increase linked to cognitive decline and underlying morphological changes	[52]
2017	20 SCZ 20 CTL	PBMCs	Individual	Treated	MALDI-TOF MS	-	No/No	MS		Alpha defensins ↑;	Activation of immune pathway of PBMCs (suggestion)	[53]
2017	60 SCZ 77 CTL	Serum	Individual	Drug naive	LC-MS/MS-MRM	-	No/No	MS	77 proteins (68 analyzed after QC) quantified of a total of 101 selected proteins	HPT ↑; ICI ↑; ANT3 ↑; CO4A ↑; AACT ↑; ITIH4 ↑; CO9 ↑; FCN3 ↑; A2AP ↑; APOH ↑; APOA2 ↓; APOC3 ↓; APOA4 ↓; APOC1 ↓;		[50]
2019	60 SCZ 20 CTL	Plasma	Pooled	Treated	1DE LC-MS/MS	ELISA	Yes/Yes	MS	Total detected: 10;	APOB ↑; C4A ↑;	-	[51]
2019	19 SCZ 19 BD 13 CTL	Serum	Pooled	Treated	LC-MS/MS and LC/ICP-MS	-	No/No	MS	Total ID: 11;	IGKC ↑; IGLC2 ↑; IGHG1 ↑; TRFE ↑; KV320 ↑; J3QRN2 ↑; IGHG3 ↑; KVD28 ↑; S4R460 ↑; LV325 ↑; IGHG2 ↑;	-	[62]
2019	45 SCZ 43 CTL	Plasma	Individual	Treated	1DE-LC-MS/MS	WB	No/yes	MS	1302 proteins screened and 34 selected (specific functions at CNS level); 5 proteins analyzed	BDNF ↓; GMFB ↓; RB3GAP1 ↓; (WB)	Psychoneuroimmune signaling	[56]

2019	33 SCZ 23 BD 24 CTL	Serum	Individual	Drug naïve	1DE-LC MS/MS	ELISA	Yes/No	MS	27 specific SCZ proteins	↑ A2ML1; ↑ ZN189; ↑ SMC2; ↑ FA12; ↑ ↑ AACT; ↑ APOE; ↑ A2GL; ↑ IPSP; ↑ DMD; ↑ CPN2; ↑ ABL2; ↑ ACTB; ↑ ACTG; ↑ PRKDC; ↑ DCD; ↑ RL19; ↑ LRP2; ↑ LG3BP; ↑ ITSN1; ↑ ECM1; ↑ ARMX4; ↑ ANR12; ↑ DHX29; ↑ DYH5; ↑ PINX1; ↑ CNDP1; ↑ FETUB) ↓ TNRC18; ↓ APOM; ↓ CASB; ↓ C1QA; ↓ RET4; ↓ APOD; ↓ TETN; ↓ CO8G; ↓ CO6; ↓ DESP; ↓ VGFR1; ↓ EST1; ↓ CADH5; ↓ KI67; ↓ MYT1; ↓ HORN; ↓ MAGE1; ↓ GULP1; (concentration of ANKRD12 ↑ in SCZ-elisa)	immune response, cell communication, cell growth and maintenance, protein metabolism and regulation of nucleic acid metabolism	[55]

SCZ: Schizophrenia; CTL: control; BD: Bipolar Disorder; DP: Depression; WB: Western Blot; ELISA: Enzyme-Linked Immunosorbent Assay; CNS: Central Nervous System; ‡ Despite the enrichment method used, the flow-through was also analyzed. The total of proteins identified came from both LC-MS individual analysis.
Altered proteins: entry name according to UniProt.

Main studies performed to compare SCZ vs. CTL:
SCZ vs CTL (n=13), SCZ&BD vs CTL (n=1) and SCZ vs Ctl&BD (n=1)

Considering the studies that perform the comparison between the SCZ cohort and, at least a group of CTL, it is noticed that a higher number of proteins identified as altered were found in studies that analyzed serum samples, also the most analyzed biological fluid. (see **Figure 4** and **Figure 5**). Fifteen studies aimed to explore the proteomic profile of the disorder and work on a potential identification of specific proteins that could help in the diagnosis (see **Table S1**). Considering this, 193 proteins were identified as differentially expressed between SCZ patients and clinical controls in blood-related samples (plasma, serum, and PBMCs), but none of the proteins are shared between these three blood matrices. Of the 15 studies analyzed, we can observe that there are commonly identified proteins between serum, plasma, and CSF samples (**Figure 4**). Moreover, there are proteins identified in PBMCs that were found in sweat and saliva samples (**Figure 5**). The majority of proteins that are shared between two different blood-related matrices were identified in serum and plasma samples. Of the 14 shared proteins (entry name: APOC3, ANT3, APOC1, APOF, APOC2, A2AP, F13B, FCN3, CFAB, CO4A, RET4, APOA4, APOA2, APOD), 7 are apolipoproteins. Likewise, the proteins that were identified in serum, plasma, and CSF were two apolipoproteins (entry name: APOA1, APOE). Also, there is one protein in common between serum and CSF analysis (entry name: TTHI). In the other samples analyzed, two proteins were coincident between PBMCs and saliva (entry name: DEF1 and DEF3), and one protein was identified in both PBMCs and sweat samples (entry name: G3P).
There are only 4 studies among the 19 that are not included in this comparison. Two of them do not have a CTL group and only intended to analyze the difference in plasma between patients treated and the same patients before AP therapy [48, 49]. The remaining two studies are not included since they do not share information about proteins ID [65, 66]

Note: Incongruities in proteins identification were solved and this information is explicit in Table S1.

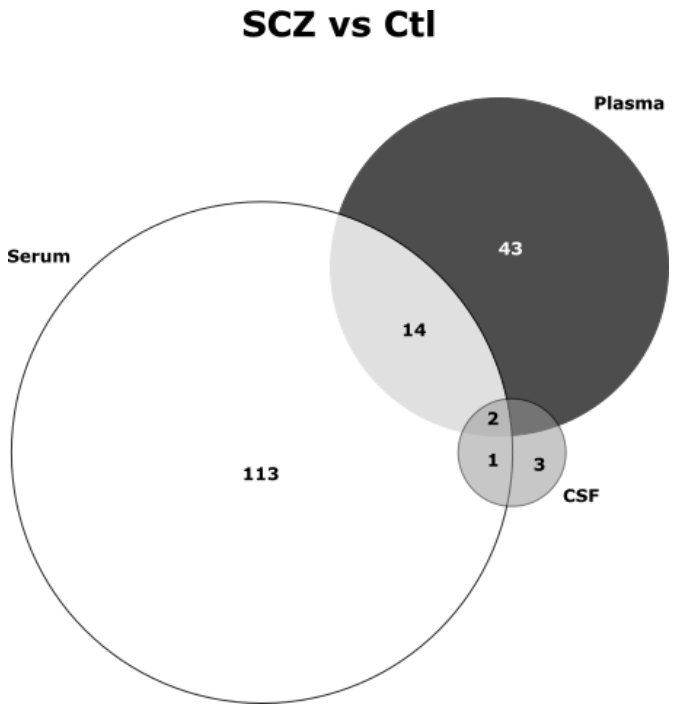


Figure 4- Venn diagram of proteins identified as altered in three biological fluids samples (serum, plasma, and CSF) in the target studies of SCZ vs. control. The proteins identified as altered in the: i) plasma vs. serum vs. CSF: 2 proteins; ii) serum vs. plasma: 14 proteins; and iii) serum vs. CSF: 1 protein. See Table S1.

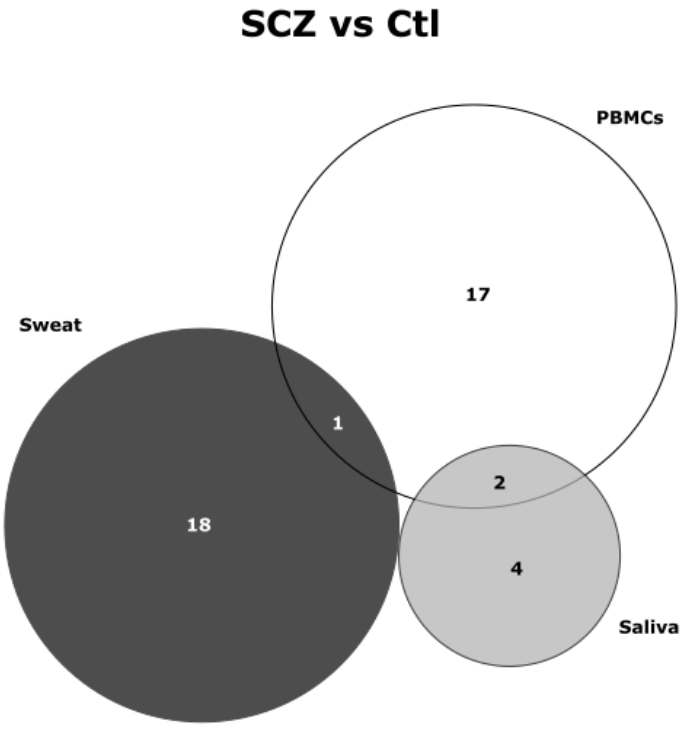


Figure 5- Venn diagram of proteins identified as altered in three biological fluids samples (sweat, PBMCs, and saliva) in the target studies of SCZ vs. control. The proteins identified as altered in the: i) PBMCs vs. saliva: 2 proteins; and ii) PBMCs vs. sweat: 1 protein. See **Table S1**.

Main studies performed to compare SCZ drug naïve vs. SCZ treated

There are 2 out of the 19 studies that aimed at comparing patients with SCZ before and after treatment (see **Table S2**). In one of these studies [49], the patients were treated for 8 weeks with the AP risperidone, and the analysis was performed before and at the end of this treatment. All patients included did not undergo prior AP treatment. This study aimed to evaluate the change in plasma protein expression levels and elucidate potential biomarkers related to metabolic side effects in consequence of risperidone treatment. Despite 18 proteins being identified as up or down-regulated after the 8 weeks of treatment, there is one protein that we did not consider in this analysis since it represents a protein isoform that we were not able to identify properly, considering the information that is available.

In another study where the comparison between a group of SCZ drug naïve/minimally medicated and the same group after treatment is performed [48], the patients were under different AP therapy: i) olanzapine (n=18); ii) quetiapine (n=14) and iii) risperidone (n=26). Patients were separated into responders (n=36) and non-responders (n=22), and the study aimed to unravel molecular pathways implicated in the efficient drug response (**Figure 6** and **Table S2**). Of the 17 proteins analyzed in the first study [49], there is only one protein that is shared with the second study and it is in common with the group of responders (entry name: NEBU).

Within the same study, where a comparison between patients who respond or do not respond to treatment is made [48], 23 identified proteins were common between both groups. Of these proteins, 13 follow the same trend: 8 proteins were down-regulated (entry names: CCN1, TBG1, ALBU, TGFI1, MLH3, APOA4, THS7A, and ZN215) while 5 proteins were up-regulated (entry names: FDX2, 5NT3B, VDAC1, TMCO2, and FARP2) in both groups. Following distinct expression, 6 proteins were down-regulated in responders and up-regulated in non-responders (entry name: U2QL1, CF157, FA9, SNX17, COR2A and MYCD), and 4 proteins were up-regulated in responders and down-regulated in non-responders (entry names: ABCAD, CCD71, ECM1, and EFCB6).

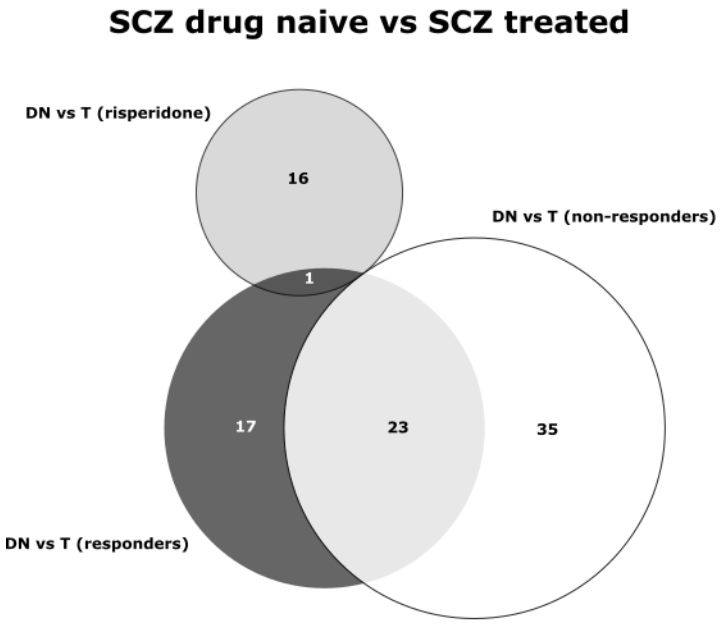


Figure 6- Venn diagram of proteins identified as altered in drug naïve patients/minimally medicated vs. treated in two studies. One circle (DN vs. T) represents one of the studies, while the two other circles (DN vs. T(responders) and DN vs. T (non-responders)) characterizes the other study. The proteins identified as altered in the: i): DN vs. T vs. DN vs. T (responders): 1 protein; and ii) DN vs. T (responders) vs. DN vs. T (non-responders): 23 proteins. See **Table S2**.

Meta-analysis observations

Of a total of 5 studies that included enough information for a meta-analysis (**Figure 7**), it could be analyzed the trend of 6 proteins (Apolipoprotein A1, Apolipoprotein A2, Apolipoprotein A4, Apolipoprotein C1, Apolipoprotein C3 and Ficolin-3) among the studies.

The meta-analysis shows a consistency of results for Apolipoprotein A1 and Apolipoprotein A2 in the 3 and 4 studies analyzed, respectively. The proteins are descreased in all studies. Nevertheless, although the uniformity showed in Apolipoprotein A1 variation, it is closed from zero (-0.29), while for Apolipoprotein A2 it is more pronounced (-0.44) . Similarly, when we analyze the effect of Apolipoprotein A4 (-0.14) and Apolipoprotein C1 (-0.23), it is also closed to the axis, with poor variation exhibited. Moreover, with an effect almost null, the studies results of Apolipoprotein A4 are not consistent, showing to be increased, decreased or close from the non-variation.

The Apolipoprotein C3 and Ficolin-3 showed the biggest effect (-0.89 and 0.78, respectively), although with different trends exhibited. Two in a total of three studies used to analyze the variation of each protein shows a consistency in results, which are decreased for Apolipoprotein C3 and increased for Ficolin-3 protein.

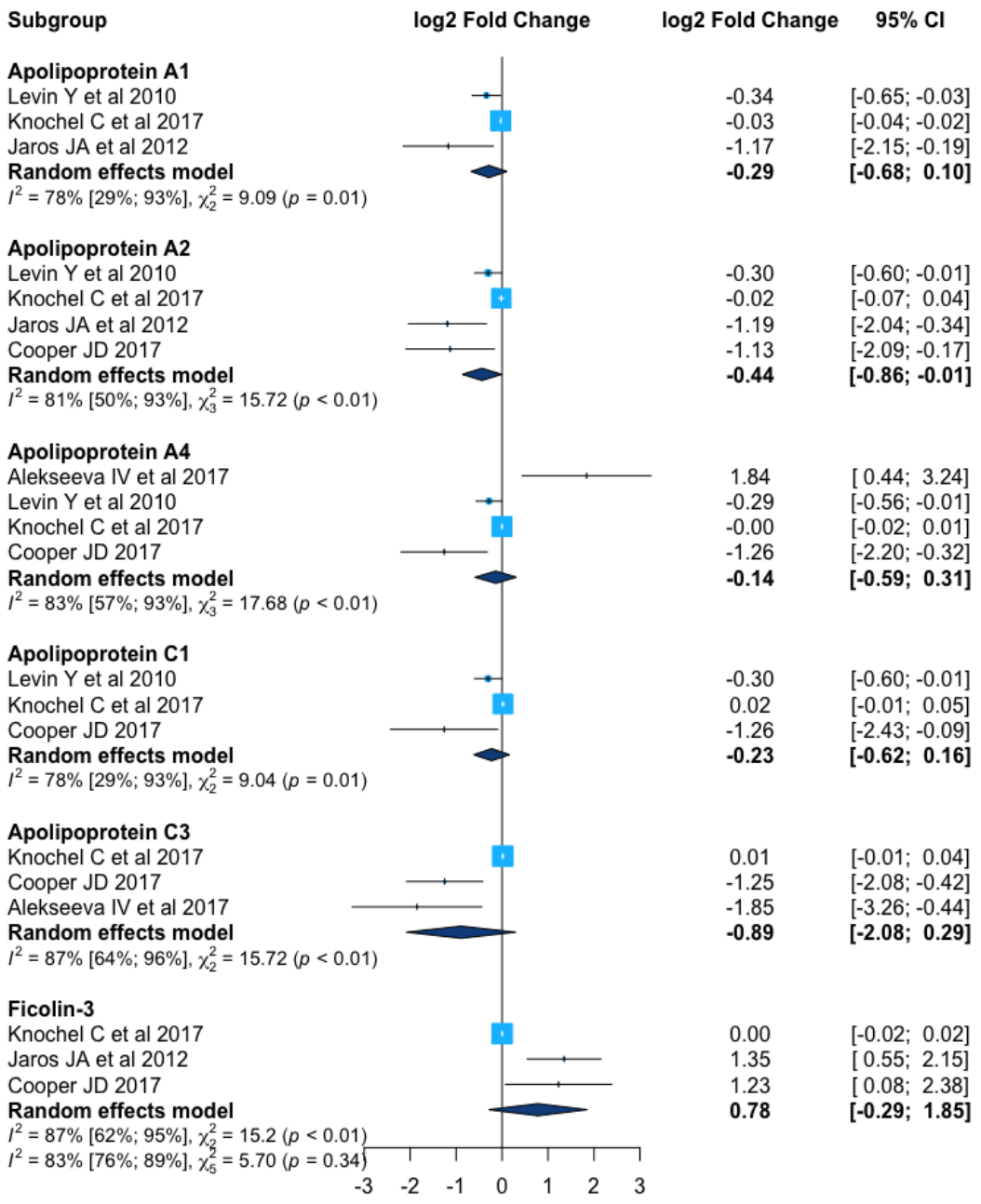


Figure 7- Forest plot showing the log2 Fold Change between SCZ and control groups' results expressed in the target studies. The overall random effects of fold change between SCZ and control groups is represented by a diamond whose width represents the 95% confidence interval (CI). I^2= index of heterogeneity; X^2 chi-squared heterogeneity statistic with degrees of freedom.

Conclusions

The present systematic review and meta-analysis revealed six proteins from the studies that passed the selection criteria. It is important to state that the information available is not uniform, which increases the difficulty to compare the studies and, in some situations, resulted in the non-inclusion of the reports in the analysis. Apolipoprotein C3 and ficolin-3 showed the bigger effects, being decreased in SCZ patients in the first case, and increased in SCZ patients in the second case. More studies with more uniform information are required to evaluate the potential use of these circulating proteins as potential biomarkers to be used in a clinical context.

Funding

This work was financed by the European Regional Development Fund (ERDF) through the COMPETE 2020 - Operational Programme for Competitiveness and Internationalisation and Portuguese national funds via FCT – Fundação para a Ciência e a Tecnologia, I.P., under projects: POCI-01-0145-FEDER-29311 (ref.: PTDC/BTM-TEC/29311/2017), PTDC/NEU-NMC/0205/2012, UIDB/ 04539/2020, POCI-01-0145-FEDER-016428 (ref.: SAICTPAC/0010/2015), POCI01-0145-FEDER-016795 (ref.: PTDC/NEU-SCC/7051/2014), POCI-01-0145-FEDER-029516 (PTDC/MED-NEU/29516/2017), POCI-01-0145-FEDER-30943 (ref.: PTDC/MEC-PSQ/30943/2017), and PTDC/MED-NEU/27946/2017; and by the National Mass Spectrometry Network (RNEM)under the contract POCI-01-0145-FEDER-402-022125 (ref.:ROTEIRO/0028/2013). CS was supported by PhD fellowship SFRH/BD/88419/2012, co-financed by the European Social Fund (ESF) through the POCH—Programa Operacional do Capital Humano and national funds via FCT.

Table S1. Proteins identified as altered in the target studies and clustered by the study comparison (SCZ vs. control, SCZ vs. BD&control, and SCZ&BD vs. control). Proteins are described by name, Unitprot and ncbi entry name (when identified as unique proteins among the studies), gene name, accession number, and the type of sample (plasma, serum, PBMCs, CSF, sweat and saliva)

Proteins identified as altered in SCZ vs. control studies				
Protein name	**Entry name**	**Gene name**	**Accsession number**	**Sample**
Alpha-2-antiplasmin	A2AP	SERPINF2	P08697	Plasma
Alpha-synuclein	E7EPV7	SNCA	E7EPV7	Plasma
Amyloid-beta precursor protein	H7C0V9	APP	P05067	Plasma
Antithrombin-III	ANT3	SERPINC1	P01008	Plasma
Apolipoprotein A1	APOA1	APOA1	P02647	Plasma
Apolipoprotein A2	APOA2	APOA2	P02652	Plasma
Apolipoprotein A4	APOA4	APOA4	P06727	Plasma
Apolipoprotein B	APOB	APOB	P04114	Plasma
Apolipoprotein B-100	APOB	APOB	P04114	Plasma
Apolipoprotein C1	APOC1	APOC1	P02654	Plasma
Apolipoprotein C2	APOC2	APOC2	P02655	Plasma
Apolipoprotein C3	APOC3	APOC3	P02656	Plasma
Apolipoprotein C4	APOC4	APOC4	P55056	Plasma
Apolipoprotein D	APOD	APOD	P05090	Plasma
Apolipoprotein E	APOE	APOE	P02649	Plasma
Apolipoprotein F	APOF	APOF	Q13790	Plasma
Apolipoprotein L1	APOL1	APOL1	O14791	Plasma
Attractin	ATRN	ATRN	O75882	Plasma
Brain-derived neurotrophic factor	BDNF	BDNF	P23560	Plasma
cAMP-dependent protein kinase type II-beta regulatory subunit	KAP3	PRKAR2B	P31323	Plasma
CDK5 regulatory subunit-associated protein 2	A0A0A0MRG9	CDK5RAP2	A0A0A0MRG9	Plasma
Clusterin	CLUS	Clu	P10909	Plasma
Coagulation factor XIII B chain	F13B	F13B	P05160	Plasma
Complement C1q subcomponent subunit C	C1QC	C1QC	P02747	Plasma
Complement C4-A	CO4A	C4A	P0C0L4	Plasma
Complement component C3	CO3	C3	P01024	Plasma
Complement factor B	CFAB	CFB	P00751	Plasma
Dihydropyrimidinase-related protein 2	DPYL2	DPYSL2	Q16555	Plasma
Drebrin	D6RFI1	DBN1	D6RFI1	Plasma
Dynamin-2	DYN2	DNM2	P50570	Plasma
Eukaryotic translation initiation factor 4 gamma 1	E7EX73	EIF4G1	E7EX73	Plasma
F-box only protein 7	FBX7	FBXO7	Q9Y3I1	Plasma

Ficolin-3	FCN3	FCN3	O75636	Plasma
Glia maturation factor beta	GMFB	GMFB	P60983	Plasma
Heat shock protein HSP 90-alpha	HS90A	HSP90AA1	P07900	Plasma
Heat shock protein HSP 90-beta	HSP90AB1	HS90B	P08238	Plasma
Heparin cofactor 2	HEP2	SERPIND1	P05546	Plasma
Histidine-rich glycoprotein	HRG	HRG	P04196	Plasma
Hypoxanthine-guanine phosphoribosyltransferase	HPRT	HPRT1	P00492	Plasma
Integral membrane protein 2B	ITM2B	ITM2B	Q9Y287	Plasma
Matrix metalloproteinase-9	MMP9	MMP9	P14780	Plasma
Maturin	MTURN	MTURN	Q8N3F0	Plasma
Mitogen-activated protein kinase 1	MK01	MAPK1	P28482	Plasma
Parkinson disease protein 7	PARK7	PARK7	Q99497	Plasma
Phospholipase A-2-activating protein	PLAP	PLAA	Q9Y263	Plasma
Pigment epithelium-derived factor	PEDF	SERPINF1	P36955	Plasma
Plasma kallikrein	KLKB1	KLKB1	P03952	Plasma
Prolow-density lipoprotein receptor-related protein 1	LRP1	LRP1	Q07954	Plasma
Rab3 GTPase-activating protein catalytic subunit	RB3GP	RAB3GAP1	Q15042	Plasma
Rab3 GTPase-activating protein non-catalytic subunit	RBGPR	RAB3GAP2	Q9H2M9	Plasma
Rap guanine nucleotide exchange factor 2	RPGF2	RAPGEF2	Q9Y4G8	Plasma
Ras-related protein Rab-11A	H3BMH2	RAB11A	H3BMH2	Plasma
Retinol-binding protein 4	RET4	RBP4	P02753	Plasma
Serine/threonine-protein kinase mTOR	MTOR	MTOR	P42345	Plasma
Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit beta isoform	E9PQN5	PPP2R5B	E9PQN5	Plasma
Signal transducer and activator of transcription	G8JLH9	STAT3	G8JLH9	Plasma
Sortilin-related receptor	SORL	SORL1	Q92673	Plasma
TBC1 domain family member 24	A0A0D9SFR5	TBC1D24	A0A0D9SFR5	Plasma
Tubulin beta-2B chain	TBB2B	TUBB2B	Q9BVA1	Plasma
Vacuolar protein sorting-associated protein 35	VPS35	VPS35	Q96QK1	Plasma
Actin, cytoplasmic 1	ACTB	ACTB	P60709	Serum
Actin, cytoplasmic 2	ACTG	ACTG1	P63261	Serum
Adhesion G protein-coupled receptor E1	AGRE1	ADGRE1	Q14246	Serum
Alpha-1-antichymotrypsin	AACT	SERPINA3	P01011	Serum
Alpha-1-antichymotrypsin	AACT	SERPINA3	P01011	Serum
Alpha-2-antiplasmin	A2AP	SERPINF2	P08697	Serum
Alpha-2-HS-glycoprotein	FETUA	AHSG	P02765	Serum
Alpha-2-macroglobulin-like protein 1	A2ML0	A2ML1	A8K2U0	Serum
Alsin	ALS2	Alsin	Q96Q42	Serum
Aminopeptidase N	AMPN	ANPEP	P15144	Serum
Ankyrin repeat domain-containing protein 12	ANR12	ANKRD12	Q6UB98	Serum

Antigen KI-67	KI67	MKI67	P46013	Serum
Antithrombin-III (ANT3)	ANT3	SERPINC1	P01008	Serum
Antithrombin-III (ANT3)	ANT3	SERPINC1	P01008	Serum
Apolipoprotein A1	APOA1	APOA1	P02647	Serum
Apolipoprotein A2	APOA2	APOA2	P02652	Serum
Apolipoprotein A4	APOA4	APOA4	P02652	Serum
Apolipoprotein A4	APOA4	APOA4	P06727	Serum
Apolipoprotein C1	APOC1	APOC1	P02654	Serum
Apolipoprotein C2 precursor	APOC2	APOC2	P02655	Serum
Apolipoprotein C3	APOC3	APOC3	P02656	Serum
Apolipoprotein D	APOD	APOD	P05090	Serum
Apolipoprotein E	APOE	APOE	P02649	Serum
Apolipoprotein F (precursor)	APOF	APOF	Q13790	Serum
Apolipoprotein M	APOM	APOM	O95445	Serum
Armadillo repeat-containing X-linked protein 4	ARMX4	ARMCX4	Q5H9R4	Serum
ATP-dependent RNA helicase	DHX29	DHX29	Q7Z478	Serum
Beta-2-glycoprotein 1	APOH	APOH	P02749	Serum
Beta-2-glycoprotein 1 (Fragment)	J3QRN2	APOH	P02749	Serum
Beta-Ala-His dipeptidase	CNDP1	CNDP1	Q96KN2	Serum
Beta-casein	CASB	CSN2	P05814	Serum
C4b-binding protein alpha chain	C4BPA	C4BPA	P04003	Serum
Cadherin-5	CADH5	CDH5	P33151	Serum
Carbonic anhydrase 1	CAH1	CA1	P00915	Serum
Carboxypeptidase N subunit 2	CPN2	CPN2	P22792	Serum
CD5 antigen-like	CD5L	CD5L	O43866	Serum
Coagulation factor VII	FA7	F7	P08709	Serum
Coagulation factor XII	FA12	F12	P00748	Serum
Coagulation factor XIII B chain	FI3B	FI3B	P05160	Serum
Coiled-coil domain-containing protein 57	CCD57	CCDC57	Q2TAC2	Serum
Complement C1q subcomponent subunit A	C1QA	C1QA	P02745	Serum
Complement C4-A	CO4A	C4A	P0C0L4	Serum
Complement component C6	CO6	C6	P13671	Serum
Complement component C8 beta chain precursor (C8 B protein)	Q05CV3	C8B	Q05CV3	Serum
Complement component C9	CO9	C9	P02748	Serum
Complement factor B	CFAB	CFB	P00751	Serum
Complement factor H-related protein 3	FHR3	CFHR3	Q02985	Serum
Dermcidin	DCD	DCD	P81605	Serum
Desmoplakin	DESP	DSP	P15924	Serum
DNA-dependent protein kinase catalytic subunit	PRKDC	PRKDC	P78527	Serum
Dopamine beta-hydroxylase	DOPO	DBH	P09172	Serum
Dynein heavy chain 5, axonemal	DNAH5	DYH5	Q8TE73	Serum
Dystrophin	DMD	DMD	P11532	Serum

Extracellular matrix protein 1	ECM1	ECM1	Q16610	Serum
Fer-1-like protein 6	FR1L6	FER1L6	Q2WGJ9	Serum
Fetuin-B	FETUB	FETUB	Q9UGM5	Serum
Fibrinogen alpha chain	FIBA	FGA	P02671	Serum
Ficolin-3	FCN3	FCN3	O75636	Serum
Galectin-3-binding protein	LG3BP	LGALS3BP	Q08380	Serum
Gamma-glutamyl hydrolase (precursor)	GGH	GGH	Q92820	Serum
Haptoglobin	HPT	HP	P00738	Serum
Hornerin	HORN	HRNR	Q86YZ3	Serum
Ig kappa chain C region (Immunoglobulin kappa constant)	IGKC	IGKC	P01834	Serum
Ig kappa chain V-III region GOL			P01619	Serum
Ig kappa chain V-III region SIE (Immunoglobulin kappa variable 3-20)	KV320	IGKV3-20	P01620	Serum
Ig lambda chain V-IV region Hil (Immunoglobulin lambda variable 3-25)	LV325	IGLV3-25	P01717	Serum
IGK@ protein	Q6PJF2	IGK@	Q6PJF2	Serum
IGL@ protein	Q8N355	IGL@	Q8N355	Serum
Immunoglobulin heavy constant gamma 1	IGHG1	IGHG1	P01857	Serum
Immunoglobulin heavy constant gamma 2	IGHG2	IGHG2	P01859	Serum
Immunoglobulin heavy constant gamma 3	IGHG3	IGHG3	P01860	Serum
Immunoglobulin heavy constant gamma 4 protein	IGHG4	IGHG4	P01861	Serum
Immunoglobulin heavy constant mu	IGHM	IGHM	P01871	Serum
Immunoglobulin lambda constant 2	IGLC2	IGLC2	P0DOY2	Serum
Insulin-like growth fac.-bind. prot. 3	IBP3	IGFBP3	P17936	Serum
Insulin-like growth factor-binding protein complex acid labile subunit (chain precursor)	ALS	IGFALS	P35858	Serum
Inter-alpha-trypsin inhibitor heavy chain H3	ITIH3	ITIH3	Q06033	Serum
Inter-alpha-trypsin inhibitor heavy chain H4	ITIH4	ITIH4	Q14624	Serum
Intercellular adhesion molecule 2 (precursor)	ICAM2	ICAM2	P13598	Serum
Intersectin-1	ITSN1	ITSN1	Q15811	Serum
Isoform 1 of C4b-binding protein beta chain precursor (C4b-binding protein beta chain)	C4BPB	C4BPB	P20851	Serum
Isoform 2 of Apolipoprotein L1 precursor (Apolipoprotein L1)	APOL1	APOL1	O14791-2	Serum
Isoform 2 of Inter-alpha-trypsin inhibitor heavy chain H4 precursor	ITIH4	ITIH4	Q14624	Serum
Keratin, type I cytoskeletal 14	K1C14	KRT14	P02533	Serum
Keratin, type I cytoskeletal 9	K1C9	KRT9	P35527	Serum
Keratin, type II cytoskeletal 6B	K2C6B	KRT6B	P04259	Serum
Kinesin-like protein KIF21B	KI21B	KIF21B	O75037	Serum
Kinesin-like protein KIF27	KIF27	KIF27	Q86VH2	Serum

Kininogen-1 (Isoform HMW of Kininogen-1 precursor)	KNG1	KNG1	P01042	Serum
Leucine-rich alpha-2-glycoprotein	A2GL	LRG1	P02750	Serum
Leucine-rich repeat-containing protein 7	LRRC7	LRRC7	Q96NW7	Serum
Leukemia inhibitory factor receptor	LIFR	LIFR	P42702	Serum
Lipopolysaccharide-binding protein (precursor)	LBP	LBP	P18428	Serum
Liver carboxylesterase 1	EST1	CES1	P23141	Serum
Low-density lipoprotein receptor-related protein 2	LRP2	LRP2	P98164	Serum
Melanoma-associated antigen E1	MAGE1	MAGEE1	Q9HCI5	Serum
Microtubule-associated serine/threonine-protein kinase 1	MAST1	MAST1	Q9Y2H9	Serum
Myelin transcription factor 1	MYT1	MYT1	Q01538	Serum
Myoferlin	MYOF	MYOF	Q9NZM1	Serum
N-acetylmuramoyl-L-alanine amidase (Isoform 1 of N-acetylmuramoyl-L-alanine amidase precursor)	PGRP2	PGLYRP2	Q96PD5	Serum
Nucleobindin-1 (CALNUC)	NUCB1	NUCB1	Q02818	Serum
Nucleoside diphos.-linked moiety X mot.6	NUDT6	NUDT6	P53370	Serum
Peptidase inhibitor 16 (Isoform 1 of Peptidase inhibitor 16 precursor)	PI16	PI16	Q6UXB8	Serum
Phosphatidylinositol-glycan-specific phospholipase D	PHLD	GPLD1	P80108	Serum
Phospholipid transfer protein (Isoform 1 of Phospholipid transfer protein precursor)	PLTP	PLTP	P55058	Serum
PIN2/TERF1-interacting telomerase inhibitor 1	PINX1	PINX1	Q96BK5	Serum
Plasma protease C1 inhibitor	IC1	SERPING1	P05155	Serum
Plasma protease C1 inhibitor	IC1	SERPING1	P05155	Serum
Plasma serine protease inhibitor	IPSP	SERPINA5	P05154	Serum
Protein IGHV3OR16-9	S4R460	IGKV2D-28	S4R460	Serum
Protein IGKV2D-28 (Immunoglobulin kappa variable 2D-28)	KVD28		A0A0A0MTQ6	Serum
PTB domain-containing engulfment adapter protein 1	GULP1	GULP1	Q9UBP9	Serum
Retinol-binding protein 4	RET4	RBP4	P02753	Serum
Ribosomal protein L19	RL19	RPL19	P84098	Serum
Sarcoma antigen 1	SAGE1	SAGE1	Q9NXZ1	Serum
Serotransferrin	TRFE	TF	P02787	Serum
Serum amyloid A1 protein preproprotein	SAA1	SAA1	gi|40316910	Serum
Sorbin and SH3 domain-containing prot.1	SRBS1	SORBS1	Q9BX66	Serum
SP 40, 40, partial/clusterin	CLUS	CLU	gi|338305	Serum
Structural maintenance of chromosomes protein 1A	SMC1A	SMC1A	Q14683	Serum
Structural maintenance of chromosomes protein 2	SMC2	SMC2	O95347	Serum
Structural maintenance of chromosomes protein 4	SMC4	SMC4	Q9NTJ3	Serum

Tetranectin	TETN	CLEC3B	P05452	Serum
Transthyretin (dimer)	TTHI	TTR	gi|114318993	Serum
Transthyretin precursor	TTHI	TTR	P02766	Serum
Trinucleotide repeat-containing gene 18 protein	TNC18	TNRC18	O15417	Serum
Tyrosine-protein kinase ABL2 (Abelson tyrosine-protein kinase 2)	ABL2	ABL2	P42684	Serum
Vascular endothelial growth factor receptor 1	VGFR1	FLT1	P17948	Serum
Vitamin K-dependent protein S (precursor)	PROS	PROS1	P07225	Serum
WD repeat-containing protein 19	WDR19	WDR19	Q8NEZ3	Serum
Xin actin-bind. repeat-containing protein 1	XIRP1	XIRP1	Q702N8	Serum
Zinc finger protein 189	ZNF189	ZNF189	O75820	Serum
Zinc finger protein 57	ZNF57	ZNF57	Q68EA5	Serum
60 kDa heat shock protein mitochondrial	CH60	HSPD1	P10809	PBMCs
alpha defensins 1	DEF1	DEFA1	P59665	PBMCs
alpha defensins 3	DEF3	DEFA3	P59666	PBMCs
Coactosin-like protein	COTL1	COTL1	Q14019	PBMCs
Cytosolic non-specific dipeptidase	CNDP2	CNDP2	Q96KP4	PBMCs
Fructose bisphosphate aldolase C	ALDOC	ALDOC	P09972	PBMCs
Glucose-6-phosphate isomerase	G6PI	GPI	P06744	PBMCs
Glyceraldehyde-3-phosphate dehydrogenase	G3P	GAPDH	P04406	PBMCs
Heat shock 70 kDa protein	HSP72	HSOA2	P54652	PBMCs
Heterogeneous nuclear ribonucleoprotein K	HNRPK	HNRNPK	P61978	PBMCs
L-lactate dehydrogenase B chain	LDHB	LDHB	P07195	PBMCs
Myosin 14	MYH14	MYH14	Q7Z406	PBMCs
Myosin 15	MYH15	MYH15	Q9Y2K3	PBMCs
Nicotinamide phosphoribosyltransferase	NAMPT	NAMPT	P43490	PBMCs
Peptidyl-prolyl cis-trans isomerase A (cyclophilin A)	PPIA	PPIA	P62937	PBMCs
Phosphoglycerate kinase 1	PGK1	PGK1	P00558	PBMCs
Probable phosphoglycerate mutase 4	PGAM4	PGAM4	Q8N0Y7	PBMCs
Pyruvate kinase PKLR	KPYR	PKLR	P30613	PBMCs
TOG array regulator of axonemal microtubules protein 1	TGRM1	TOGARAM1	Q9Y4F4	PBMCs
Triosephosphate isomerase	TPIS	TPI1	P60174	PBMCs
Apolipoprotein A1	APOA1	APOA1	P02647	CSF
Apolipoprotein E	APOE	APOE	P02649	CSF
Coiled-coil domain-containing protein 3	CCDC3	CCDC3	Q9BQI4	CSF
Prostaglandin-H2 D-isomerase	PTGDS	PTGDS	P41222	CSF
TGF-β receptor	TGFR1	TGFBR1	P36897	CSF
Transthyretin precursor	TTHY	TTR	P02766	CSF
annexin A5	ANXA5	ANXA5	gi|4502107	Sweat
arginase-1 isoform 2	ARGI1	ARG1	gi|10947139	Sweat

bleomycin hydrolase	BLMH	BLMH	gi|4557367	Sweat
calmodulin-like skin protein	CALL5	CALML5	gi|8393159	Sweat
caspase 14 precursor	CASPE	CASP14	gi|6912286	Sweat
Corneodesmosin	CDSN	CDSN	gi|67782356	Sweat
Cystatin A	CYTA	CSTA	gi|4885165|	Sweat
dermcidin isoform 1 preproprotein	DCD	DCD	gi|16751921	Sweat
desmoglein 1 preproprotein	DSG1	DSG1	gi|4503401	Sweat
glyceraldehyde-3-phosphate dehydrogenase	G3P	GAPDH	P04406	Sweat
kallikrein 11 isoform 1 precursor	KLK11	KLK11	gi|5803199	Sweat
keratin, type I cytoskeletal 10	K1C10	KRT10	gi|40354192	Sweat
peroxiredoxin-1	PRDX1	PRDX1	gi|32455264	Sweat
phosphatidylethanolamine-binding protein 1	PEBP1	PEBP1	gi|4505621	Sweat
prolactin-inducible protein precursor	PIP	PIP	gi|4505821	Sweat
protein S100-A7	S10A7	S100A7	gi|4506769	Sweat
protein/nucleic acid deglycase DJ-1	PARK7	PARK7	gi|31543380	Sweat
thioredoxin isoform 1	THIO	TXN	gi|50592994	Sweat
zinc-alpha-2-glycoprotein precursor	ZA2G	AZGP1	gi|4502337	Sweat
alpha defensins 1	DEF1	DEFA1	P59665	Saliva
alpha defensins 3	DEF3	DEFA3	P59666	Saliva
alpha defensins 4	DEF4	DEFA4	P12838	Saliva
cystatin A	CYTA	CSTA	P01040	Saliva
Cystatin B S-cysteinyl	CYTB	CSTB	P04080	Saliva
S100A12	S10AC	S100A12	P80511	Saliva

Note: Of the proteins identified as altered in three biological fluids samples (serum, plasma, and CSF), 1 protein (alpha defensins 2) identified in saliva and PBMCs was not considered since it was not given information about protein ID and we were not able to find the corresponding accession number/identifier through UniProt database [53, 58]; Albumin fragment and alpha2-antitrypsin were identified, each protein, in one serum study but the accession number does not correspond to the reported protein; In the same study, an Apolipoprotein C2 was wrongly matched with the entry name, gene and accession number and, for this reason, was also excluded from this comparison; Likewise, 3 identified Haptoglobin hp2α in this study were not considered for this analysis [54]; In another serum study, one protein had its accession number updated, taking into consideration UniProt data (from P04206 to P01619) [60];

Of the proteins identified as altered in three biological fluids samples (sweat, PBMCs, and saliva), only a protein isoform from the study that analyzed individuals with SCZ after risperidone treatment was not considered. Without the accession number information, we could not identify this protein properly to take into consideration for this analysis [49].

Table S2. Proteins identified as altered in the target studies and clustered by the study comparison (SCZ vs. treated). Proteins are described by name, Unitprot and ncbi entry name (when identified as unique proteins among the studies), gene name, accession number, and the cohort used/group of study (DN vs T (risperidone) and DN vs T (responders and non-responders)).

Proteins identified as altered in SCZ vs. SCZ treated				
Protein name	**Entry name**	**Gene name**	**Accsession number**	**Group of study**
(Interferon-induced) guanylate-binding protein 1	GBP1	GBP1	Q01514	DN vs T (risperidone)
(Isoform1 of) zinc finger protein185	ZN185	ZNF185	O15231	DN vs T (risperidone)
Apolipoprotein A1	APOA1	APOA1	P02647	DN vs T (risperidone)
Complement component 4B preproprotein	H9YW54	C4B	H9YW54	DN vs T (risperidone)
Complement component C8 beta chain	CO8B	C8B	P07358	DN vs T (risperidone)
Complement factor B	CFAB	CFB	P00751	DN vs T (risperidone)
Gelsolin	GELS	GSN	P06396	DN vs T (risperidone)
Hemopexin	HEMO	HPX	P02790	DN vs T (risperidone)
Histamine N-methyltransferase	HNMT	HNMT	P50135	DN vs T (risperidone)
Isoform2 of vinculin	VINC	VCL	P18206	DN vs T (risperidone)
Keratin, type I cytoskeletal 9	K1C9	KRT9	P35527	DN vs T (risperidone)
Keratin, type II cytoskeletal 1	K2C1	KRT1	P04264	DN vs T (risperidone)
Nebulin	NEBU	NEB	P20929	DN vs T (risperidone)
Plasminogen	PLMN	PLG	P00747	DN vs T (risperidone)
Protein ALEX	ALEX	GNAS	P84996	DN vs T (risperidone)
retinol-binding protein4	RET4	RBP4	P02753	DN vs T (risperidone)
Serotransferrin	TRFE	TF	P02787	DN vs T (risperidone)
39S ribosomal protein L3, mitochondrial	RM03	MRPL3	P09001	DN vs T (responders)
7-methylguanosine phosphate-specific 5'-nucleotidase	5NT3B	NT5C3B	Q969T7	DN vs T (responders)
Alpha-1-antichymotrypsin	AACT	SERPINA3	P01011	DN vs T (responders)
Apolipoprotein A4	APOA4	APOA4	P06727	DN vs T (responders)
ATP-binding cassette sub-family A member 13	ABCAD	ABCA13	Q86UQ4	DN vs T (responders)
ATP-binding cassette sub-family B member 5	ABCB5	ABCB5	Q2M3G0	DN vs T (responders)
Capping protein, Arp2/3 and myosin-I linker protein 2	CARL2	CARMIL2	Q6F5E8	DN vs T (responders)
CCN family member 1	CCN1	CCN1	O00622	DN vs T (responders)

CD5 antigen-like	CD5L	CD5L	O43866	DN vs T (responders)
Centromere protein P	CENPP	CENPP	Q6IPU0	DN vs T (responders)
Cilia- and flagella-associated protein 157	CF157	CFAP157	Q5JU67	DN vs T (responders)
Coagulation factor IX	FA9	F9	P00740	DN vs T (responders)
Coiled-coil domain-containing protein 71	CCD71	CCDC71	Q8IV32	DN vs T (responders)
Coiled-coil domain-containing protein 74A	CC74A	CCDC74A	Q96AQ1	DN vs T (responders)
Collagen alpha-1(XXV) chain	COPA1	COL25A1	Q9BXS0	DN vs T (responders)
Coronin-2A	COR2A	CORO2A	Q92828	DN vs T (responders)
DNA mismatch repair protein Mlh3	MLH3	MLH3	Q9UHC1	DN vs T (responders)
DNA-dependent protein kinase catalytic subunit	PRKDC	PRKDC	P78527	DN vs T (responders)
E3 ubiquitin-protein ligase MYCBP2	MYCB2	MYCBP2	O75592	DN vs T (responders)
EF-hand calcium-binding domaincontaining protein 6	EFCB6	EFCAB6	Q5THR3	DN vs T (responders)
Engulfment and cell motility protein 1	ELMO1	ELMO1	Q92556	DN vs T (responders)
Extracellular matrix protein 1	ECM1	ECM1	Q16610	DN vs T (responders)
FERM, RhoGEF and pleckstrin domaincontaining protein 2	FARP2	FARP2	O94887	DN vs T (responders)
Ferredoxin-2, mitochondrial	FDX2	FDX2	Q6P4F2	DN vs T (responders)
Heat shock 70 kDa protein 1-like	HS71L_HUMAN	HSPA1L	P34931	DN vs T (responders)
Inter-alpha-trypsin inhibitor heavy chain H3	ITIH3	ITIH3	Q06033	DN vs T (responders)
Myocardin	MYCD	MYOCD	Q8IZQ8	DN vs T (responders)
Nebulin	NEBU	NEB	P20929	DN vs T (responders)
PDZ domain-containing protein 2	PDZD2	PDZD2	O15018	DN vs T (responders)
Protein limb expression 1 homolog	LIX1	LIX1	Q8N485	DN vs T (responders)
Serum albumin	ALBU	ALB	P02768	DN vs T (responders)
Sorting nexin-17	SNX17	SNX17	Q15036	DN vs T (responders)
Thrombospondin type-1 domaincontaining protein 7A	THS7A	THSD7A	Q9UPZ6	DN vs T (responders)
Transforming growth factor beta-1-induced transcript 1 protein	TGFI1	TGFB1I1	O43294	DN vs T (responders)
Transmembrane and coiled-coil domain-containing protein 2	TMCO2	TMCO2	Q7Z6W1	DN vs T (responders)
Tubulin gamma-1 chain	TBG1	TUBG1	P23258	DN vs T (responders)
Tumor protein p73	P73	TP73	O15350	DN vs T (responders)

Ubiquitin-conjugating enzyme E2Q-like protein 1	U2QL1	UBE2QL1	A1L167	DN vs T (responders)
Uncharacterized protein HSD47	YO003	HSD47?	Q6ZVN6	DN vs T (responders)
Voltage-dependent anion-selective channel protein 1	VDAC1	VDAC1	P21796	DN vs T (responders)
Zinc finger protein 215	ZN215	ZNF215	Q9UL58	DN vs T (responders)
7-methylguanosine phosphatespecific 5'-nucleotidase	5NT3B	NT5C3B	Q969T7	DN vs T (non-responders)
Apolipoprotein A4	APOA4	APOA4	P06727	DN vs T (non-responders)
Apolipoprotein B-100	APOB	APOB	P04114	DN vs T (non-responders)
ATP-binding cassette sub-family A member 13	ABCAD	ABCA13	Q86UQ4	DN vs T (non-responders)
ATP-dependent Clp protease ATP-binding subunit clpX-like, mitochondrial	CLPX	CLPX	O76031	DN vs T (non-responders)
cAMP-dependent protein kinase catalytic subunit PRKX	PRKX	PRKX	P51817	DN vs T (non-responders)
CCN family member 1	CCN1	CCN1	O00622	DN vs T (non-responders)
Chromogranin-A	CMGA	CHGA	P10645	DN vs T (non-responders)
Cilia- and flagella-associated protein 157	CF157	CFAP157	Q5JU67	DN vs T (non-responders)
Coagulation factor IX	FA9	F9	P00740	DN vs T (non-responders)
Coiled-coil domain-containing protein 71	CCD71	CCDC71	Q8IV32	DN vs T (non-responders)
Complement C2	CO2	C2	P06681	DN vs T (non-responders)
Coronin-2A	COR2A	CORO2A	Q92828	DN vs T (non-responders)
DNA (cytosine-5)-methyltransferase 1	DNMT1	DNMT1	P26358	DN vs T (non-responders)
DNA mismatch repair protein Mlh3	MLH3	MLH3	Q9UHC1	DN vs T (non-responders)
E3 ubiquitin-protein ligase Arkadia	RNF11	RNF11	Q6ZNA4	DN vs T (non-responders)
EF-hand calcium-binding domaincontaining protein 6	EFCB6	EFCAB6	Q5THR3	DN vs T (non-responders)

Epithelial cell-transforming sequence 2 oncogene-like	ECT2L	ECT2L	Q008S8	DN vs T (non-responders)
Extracellular matrix protein 1	ECM1	ECM1	Q16610	DN vs T (non-responders)
FERM, RhoGEF and pleckstrin domain-containing protein 2	FARP2	FARP2	O94887	DN vs T (non-responders)
Ferredoxin-2, mitochondrial	FDX2	FDX2	Q6P4F2	DN vs T (non-responders)
Glycerophosphodiester phosphodiesterase domaincontaining protein 5	GDPD5	GDPD5	Q8WTR4	DN vs T (non-responders)
Golgin sub-family A member 4	GOGA4	GOLGA4	Q13439	DN vs T (non-responders)
Histone demethylase UTY	UTY	UTY	O14607	DN vs T (non-responders)
Inactive serine/threonine-protein kinase TEX14	TEX14	TEX14	Q8IWB6	DN vs T (non-responders)
Kinesin-like protein KIF17	KIF17	KIF17	Q9P2E2	DN vs T (non-responders)
Kinesin-like protein KIF3A	KIF3A	KIF3A	Q9Y496	DN vs T (non-responders)
Major vault protein	MVP	MVP	Q14764	DN vs T (non-responders)
Mucin-16	MUC16	MUC16	Q8WXI7	DN vs T (non-responders)
Myocardin	MYCD	MYOCD	Q8IZQ8	DN vs T (non-responders)
Myosin-binding protein C, slow-type	MYPC1	MYBPC1	Q00872	DN vs T (non-responders)
Olfactory receptor 4K13	OR4KD	OR4K13	Q8NH42	DN vs T (non-responders)
Poly [ADP-ribose] polymerase 2	PARP2	PARP2	Q9UGN5	DN vs T (non-responders)
Poly(rC)-binding protein 2	PCBP2	PCBP2	Q15366	DN vs T (non-responders)
Probable ATP-dependent RNA helicase DDX46	DDX46	DDX46	Q7L014	DN vs T (non-responders)
Protein SPT2 homolog	SPT2	SPTY2D1	Q68D10	DN vs T (non-responders)
Putative fidgetin-like protein 2	FIGL2	FIGNL2	A6NMB9	DN vs T (non-responders)

SEC14-like protein 4	S14L4	SEC14L4	Q9UDX3	DN vs T (non-responders)
Septin-7	SEPT7	SEPT07	Q16181	DN vs T (non-responders)
Serum albumin	ALBU	ALB	P02768	DN vs T (non-responders)
Serum amyloid P-component	SAMP	APCS	P02743	DN vs T (non-responders)
Sorting nexin-17	SNX17	SNX17	Q15036	DN vs T (non-responders)
Synapsin-3	SYN3	SYN3	O14994	DN vs T (non-responders)
Thrombospondin type-1 domaincontaining protein 7A	THS7A	THSD7A	Q9UPZ6	DN vs T (non-responders)
Titin	TITIN	TTN	Q8WZ42	DN vs T (non-responders)
Transforming growth factor beta-1-induced transcript 1 protein	TGFI1	TGFB1I1	O43294	DN vs T (non-responders)
Transmembrane and coiled-coil domain-containing protein 2	TMCO2	TMCO2	Q7Z6W1	DN vs T (non-responders)
Transmembrane protein KIAA1109	K1109	KIAA1109	Q2LD37	DN vs T (non-responders)
Tubulin gamma-1 chain	TBG1	TUBG1	P23258	DN vs T (non-responders)
Ubiquitin-conjugating enzyme E2Q-like protein 1	U2QL1	UBE2QL1	A1L167	DN vs T (non-responders)
Vesicular, overexpressed in cancer, prosurvival protein 1	VOPP1	VOPP1	Q96AW1	DN vs T (non-responders)
Vitamin K-dependent protein S	PROS	PROS1	P07225	DN vs T (non-responders)
Voltage-dependent anion-selective channel protein 1	VDAC1	VDAC1	P21796	DN vs T (non-responders)
WD repeat-containing protein 11	WDR11	WDR11	Q9BZH6	DN vs T (non-responders)
Wee1-like protein kinase	WEE1	WEE1	P30291	DN vs T (non-responders)
Zinc finger CCCH domain-containing protein 13	ZC3HD	ZC3H13	Q5T200	DN vs T (non-responders)
Zinc finger homeobox protein 4	ZFHX4	ZFHX4	Q86UP3	DN vs T (non-responders)

Zinc finger protein 215	ZN215	ZNF215	Q9UL58	DN vs T (non-responders)

References

1. Hyman, S.E., *A glimmer of light for neuropsychiatric disorders.* Nature, 2008. **455**(7215): p. 890-3.
2. Peedicayil, J., *Identification of Biomarkers in Neuropsychiatric Disorders Based on Systems Biology and Epigenetics.* Front Genet, 2019. **10**: p. 985.
3. Guest, F.L., P.C. Guest, and D. Martins-de-Souza, *The emergence of point-of-care blood-based biomarker testing for psychiatric disorders: enabling personalized medicine.* Biomark Med, 2016. **10**(4): p. 431-43.
4. Taurines, R., et al., *Proteomic research in psychiatry.* J Psychopharmacol, 2011. **25**(2): p. 151-96.
5. Hawton, K., *van HK. Suicide.* Lancet, 2009. **373**(9672): p. 1372-81.
6. Preece, R.L., S.Y.S. Han, and S. Bahn, *Proteomic approaches to identify blood-based biomarkers for depression and bipolar disorders.* Expert Review of Proteomics, 2018. **15**(4): p. 325-340.
7. Trautmann, S., J. Rehm, and H.U. Wittchen, *The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders?* EMBO Rep, 2016. **17**(9): p. 1245-9.
8. Zuccoli, G.S., et al., *The Energy Metabolism Dysfunction in Psychiatric Disorders Postmortem Brains: Focus on Proteomic Evidence.* Frontiers in Neuroscience, 2017. **11**: p. 14.
9. Nestler, E.J. and S.E. Hyman, *Animal models of neuropsychiatric disorders.* Nat Neurosci, 2010. **13**(10): p. 1161-9.
10. Schizophrenia Working Group of the Psychiatric Genomics, C., *Biological insights from 108 schizophrenia-associated genetic loci.* Nature, 2014. **511**(7510): p. 421-7.
11. Gulsuner, S., et al., *Genetics of schizophrenia in the South African Xhosa.* Science, 2020. **367**(6477): p. 569-573.
12. Charlson, F.J., et al., *Global Epidemiology and Burden of Schizophrenia: Findings From the Global Burden of Disease Study 2016.* Schizophr Bull, 2018. **44**(6): p. 1195-1203.
13. Giusti, L., et al., *Proteomics insight into psychiatric disorders: an update on biological fluid biomarkers.* Expert Rev Proteomics, 2016. **13**(10): p. 941-950.
14. Alawam, K., *Application of proteomics in diagnosis of ADHD, schizophrenia, major depression, and suicidal behavior.* Adv Protein Chem Struct Biol, 2014. **95**: p. 283-315.
15. Cohen, C.I., P.D. Meesters, and J. Zhao, *New perspectives on schizophrenia in later life: implications for treatment, policy, and research.* The Lancet Psychiatry, 2015. **2**(4): p. 340-350.
16. Aricioglu, F., et al., *Neuroinflammation in Schizophrenia: A Critical Review and The Future.* Klinik Psikofarmakoloji Bülteni-Bulletin of Clinical Psychopharmacology, 2017. **26**(4): p. 429-437.
17. Davison, J., et al., *A systematic review of metabolite biomarkers of schizophrenia.* Schizophr Res, 2018. **195**: p. 32-50.
18. Santa, C., et al., *Circulating biomarkers in schizophrenia: a proteomics perspective.* International Journal of Clinical Neurosciences and Mental Health, 2017(4(Suppl. 3)): p. S05.
19. Taber, K.H., R.A. Hurley, and S.C. Yudofsky, *Diagnosis and treatment of neuropsychiatric disorders.* Annu Rev Med, 2010. **61**: p. 121-33.
20. McGorry, P.D., et al., *Back to the future: predicting and reshaping the course of psychotic disorder.* Archives of General Psychiatry, 2008. **65**(1): p. 25-27.
21. Owen, M.J., A. Sawa, and P.B. Mortensen, *Schizophrenia.* The Lancet, 2016. **388**(10039): p. 86-97.
22. Kumar, A., et al., *Potential drug targets and treatment of schizophrenia.* Inflammopharmacology, 2017. **25**(3): p. 277-292.

23. Lally, J. and J.H. MacCabe, *Antipsychotic medication in schizophrenia: a review.* Br Med Bull, 2015. **114**(1): p. 169-79.
24. Nascimento, J.M., et al., *Proteomics and molecular tools for unveiling missing links in the biochemical understanding of schizophrenia.* Proteomics Clin Appl, 2016. **10**(12): p. 1148-1158.
25. Brisch, R., et al., *The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: old fashioned, but still in vogue.* Front Psychiatry, 2014. **5**: p. 47.
26. Girgis, R.R., S.S. Kumar, and A.S. Brown, *The cytokine model of schizophrenia: emerging therapeutic strategies.* Biol Psychiatry, 2014. **75**(4): p. 292-299.
27. Pantelis, C., et al., *Structural brain imaging evidence for multiple pathological processes at different stages of brain development in schizophrenia.* Schizophr Bull, 2005. **31**(3): p. 672-96.
28. Filiou, M.D. and C.W. Turck, *General overview: biomarkers in neuroscience research.* Int Rev Neurobiol, 2011. **101**: p. 1-17.
29. Canuel, V., et al., *Translational research platforms integrating clinical and omics data: a review of publicly available solutions.* Brief Bioinform, 2015. **16**(2): p. 280-90.
30. McShane, L.M., et al., *Criteria for the use of omics-based predictors in clinical trials.* Nature, 2013. **502**(7471): p. 317-20.
31. Biomarkers Definitions Working, G., *Biomarkers and surrogate endpoints: preferred definitions and conceptual framework.* Clin Pharmacol Ther, 2001. **69**(3): p. 89-95.
32. Bayes, A. and S.G. Grant, *Neuroproteomics: understanding the molecular organization and complexity of the brain.* Nat Rev Neurosci, 2009. **10**(9): p. 635-46.
33. Kitchen, R.R., et al., *Decoding neuroproteomics: integrating the genome, translatome and functional anatomy.* Nat Neurosci, 2014. **17**(11): p. 1491-9.
34. Santa, C., et al., *Neuroproteomics — LC-MS Quantitative Approaches.* 2015.
35. Gottschalk, M.G., et al., *Proteomic enrichment analysis of psychotic and affective disorders reveals common signatures in presynaptic glutamatergic signaling and energy metabolism.* Int J Neuropsychopharmacol, 2014. **18**(2).
36. de Jesus, J.R., et al., *Bipolar disorder: recent advances and future trends in bioanalytical developments for biomarker discovery.* Anal Bioanal Chem, 2015. **407**(3): p. 661-7.
37. Chan, M.K., et al., *Converging evidence of blood-based biomarkers for schizophrenia: an update.* Int Rev Neurobiol, 2011. **101**: p. 95-144.
38. de Jesus, J.R., et al., *Bipolar disorder: recent advances and future trends in bioanalytical developments for biomarker discovery.* Analytical and Bioanalytical Chemistry, 2015. **407**(3): p. 661-667.
39. Guest, P.C., et al., *The use of proteomic biomarkers for improved diagnosis and stratification of schizophrenia patients.* Biomark Med, 2014. **8**(1): p. 15-27.
40. Cunningham, R., D. Ma, and L. Li, *Mass Spectrometry-based Proteomics and Peptidomics for Systems Biology and Biomarker Discovery.* Front Biol (Beijing), 2012. **7**(4): p. 313-335.
41. Wei, X. and L. Li, *Mass spectrometry-based proteomics and peptidomics for biomarker discovery in neurodegenerative diseases.* Int J Clin Exp Pathol, 2009. **2**(2): p. 132-48.
42. Chandramouli, K. and P.Y. Qian, *Proteomics: challenges, techniques and possibilities to overcome biological sample complexity.* Hum Genomics Proteomics, 2009. **2009**.
43. Tuli, L. and H.W. Ressom, *LC-MS Based Detection of Differential Protein Expression.* J Proteomics Bioinform, 2009. **2**: p. 416-438.
44. Walker, S.J. and A. Xu, *Biomarker discovery using molecular profiling approaches*, in *International review of neurobiology*. 2004, Elsevier. p. 1-30.
45. Neilson, K.A., et al., *Less label, more free: approaches in label-free quantitative mass spectrometry.* Proteomics, 2011. **11**(4): p. 535-53.

46. Anand, S., et al., *Label-Based and Label-Free Strategies for Protein Quantitation.* Methods Mol Biol, 2017. **1549**: p. 31-43.
47. Craft, G.E., A. Chen, and A.C. Nairn, *Recent advances in quantitative neuroproteomics.* Methods, 2013. **61**(3): p. 186-218.
48. Martins-de-Souza, D., et al., *Biological pathways modulated by antipsychotics in the blood plasma of schizophrenia patients and their association to a clinical response.* NPJ Schizophr, 2015. **1**: p. 15050.
49. Song, X., et al., *APOA-I: a possible novel biomarker for metabolic side effects in first episode schizophrenia.* PLoS One, 2014. **9**(4): p. e93902.
50. 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.
51. Walss-Bass, C., et al., *X-aptamer technology identifies c4a and apob in blood as potential markers for schizophrenia.* Molecular neuropsychiatry, 2019. **5**(1): p. 52-59.
52. Knochel, 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.
53. 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.
54. Alekseeva, I.V., et al., *A distinct serum protein pattern in patients with paranoid schizophrenia.* Physiology and Pharmacology, 2017. **21**(4): p. 304-311.
55. Smirnova, L., et al., *The difference in serum proteomes in schizophrenia and bipolar disorder.* Bmc Genomics, 2019. **20**: p. 14.
56. Rodrigues-Amorim, D., et al., *Proteomics in Schizophrenia: A Gateway to Discover Potential Biomarkers of Psychoneuroimmune Pathways.* Front Psychiatry, 2019. **10**: p. 885.
57. Martins-De-Souza, D., et al., *Different apolipoprotein E, apolipoprotein A1 and prostaglandin-H2 D-isomerase levels in cerebrospinal fluid of schizophrenia patients and healthy controls.* World J Biol Psychiatry, 2010. **11**(5): p. 719-28.
58. Iavarone, F., et al., *Characterization of salivary proteins of schizophrenic and bipolar disorder patients by top-down proteomics.* Journal of Proteomics, 2014. **103**: p. 15-22.
59. Raiszadeh, M.M., et al., *Proteomic analysis of eccrine sweat: implications for the discovery of schizophrenia biomarker proteins.* J Proteome Res, 2012. **11**(4): p. 2127-39.
60. Li, Y., et al., *Label-free quantitative proteomic analysis reveals dysfunction of complement pathway in peripheral blood of schizophrenia patients: evidence for the immune hypothesis of schizophrenia.* Mol Biosyst, 2012. **8**(10): p. 2664-71.
61. Jaros, J.A., et al., *Protein phosphorylation patterns in serum from schizophrenia patients and healthy controls.* J Proteomics, 2012. **76 Spec No.**: p. 43-55.
62. de Souza Pessoa, G., et al., *Metallomics-based platforms for comparing the human blood serum profiles between bipolar disorder and schizophrenia patients.* Rapid Commun Mass Spectrom, 2019: p. e8698.
63. Levin, Y., et al., *Global proteomic profiling reveals altered proteomic signature in schizophrenia serum.* Mol Psychiatry, 2010. **15**(11): p. 1088-100.
64. Herberth, M., et al., *Impaired glycolytic response in peripheral blood mononuclear cells of first-onset antipsychotic-naive schizophrenia patients.* Mol Psychiatry, 2011. **16**(8): p. 848-59.
65. Al Awam, K., et al., *Multiplatform metabolome and proteome profiling identifies serum metabolite and protein signatures as prospective biomarkers for schizophrenia.* J Neural Transm (Vienna), 2015. **122 Suppl 1**: p. S111-22.
66. Ding, Y.H., et al., *Protein Biomarkers in Serum of Patients with Schizophrenia.* Cell Biochem Biophys, 2015. **72**(3): p. 799-805.

67. Davalieva, K., I.M. Kostovska, and A.J. Dwork, *Proteomics Research in Schizophrenia.* Frontiers in Cellular Neuroscience, 2016. **10**: p. 22.
68. Schwarz, E. and S. Bahn, *Biomarker discovery in psychiatric disorders.* Electrophoresis, 2008. **29**(13): p. 2884-90.
69. Guest, P.C., F.L. Guest, and D. Martins-de Souza, *Making Sense of Blood-Based Proteomics and Metabolomics in Psychiatric Research.* Int J Neuropsychopharmacol, 2016. **19**(6).
70. Lee, P.Y., et al., *Plasma/serum proteomics: depletion strategies for reducing high-abundance proteins for biomarker discovery.* Bioanalysis, 2019. **11**(19): p. 1799-1812.