Investigating the Transcriptome Signature of Depression: Employing Co-expression Network, Candidate Pathways and Machine Learning Approaches

A thesis submitted for the degree of Doctor of Philosophy

By

Liliana Ciobanu

Discipline of Psychiatry
Adelaide Medical School
Faculty of Health and Medical Sciences
The University of Adelaide

June 2019

Table of Contents

Publications.....3

Abstract.....4

Statement.....6

Acknowledgements.....7

Chapter 1

General introduction

Epidemiology of Major Depressive Disorder (MDD).....8

Clinical manifestation of MDD.....10

Major hypothesis of pathophysiology of MDD.....13

Why study the transcriptome in depression?.....19

Chapter 2

Genome-wide differential gene expression in brain and peripheral tissues in MDD across the life span (Ciobanu et al., 2016)

Introduction.....23

Methods.....26

Results.....28

Discussion.....42

From a single gene to a network of complex interactions.....53

Chapter 3

Weighed gene co-expression network analysis (WGCNA) of peripheral blood transcriptome in recurrent MDD in older adults

Introduction.....58

Methods.....60

Results.....64

Discussion.....72

From biological networks to machine learning predictions.....79

Chapter 4

Downregulated transferrin receptor in the blood predicts recurrent MDD in an elderly cohort: a fuzzy forests approach

Introduction 84

Methods 86

Results 88

Discussion 91

An emerging need to focus on immune dysregulations in depression 94

Chapter 5

Dysregulation of immune-related pathways in depression

Introduction 97

Methods 111

Results 114

Discussion 116

A novel candidate pathway approach 122

Chapter 6

Summarizing the findings and outlining future directions................................ 124

Gene co-expression network and machine learning approaches to translate biological findings into actionable clinical insights 127

Appendices

Appendix A 140

Appendix B 167

Appendix C 173

Bibliography 176

List of figures 195

List of table 196

Publications

During her PhD period, Liliana Ciobanu has authored and co-authored over 25 articles and two book chapters that can be accessed at:

https://www.researchgate.net/profile/Liliana_Ciobanu

The following articles are included in this thesis:

CIOBANU, L. G., SACHDEV, P. S., TROLLOR, J. N., REPPERMUND, S., THALAMUTHU, A., MATHER, K. A., COHEN-WOODS, S. & BAUNE, B. T. 2016. Differential gene expression in brain and peripheral tissues in depression across the life span: A review of replicated findings. *Neuroscience & Biobehavioral Reviews,* 71**,** 281-293.

CIOBANU, L. G., SACHDEV, P. S., TROLLOR, J. N., REPPERMUND, S., THALAMUTHU, A., MATHER, K. A., COHEN-WOODS, S., STACEY, D., TOBEN, C., SCHUBERT, K. O. & BAUNE, B. T. 2018b. Co-expression network analysis of peripheral blood transcriptome identifies dysregulated protein processing in endoplasmic reticulum and immune response in recurrent MDD in older adults. *Journal of Psychiatric Research*.

CIOBANU, L. G., SACHDEV, P. S., TROLLOR, J. N., REPPERMUND, S., THALAMUTHU, A., MATHER, K. A., COHEN-WOODS, S., STACEY, D., TOBEN, C., SCHUBERT, K. O. & BAUNE, B. T. 2019. Downregulated transferrin receptor in the blood predicts recurrent MDD in the elderly cohort: a Fuzzy Forests approach. *Journal of Psychiatric Research*. Under review

The parts of the following book chapters are included in this thesis:

CIOBANU, L. G. & BAUNE, B. T. 2018. Chapter 11 - Gene Expression of Inflammation Markers in Depression. *In:* BAUNE, B. T. (ed.) *Inflammation and Immunity in Depression.* Academic Press.

CIOBANU, L. G. & BAUNE, B. T. 2019. Gene co-expression network and machine learning in personalized psychiatry. *In:* BAUNE, B. T. (ed.) *Personalized psychiatry. First edition.* Academic Press.

Abstract

Depression is the leading cause of disability worldwide and is one of the major contributors to the overall global burden of disease. Despite significant advances in elucidating the neurobiology of depression in recent years, the molecular factors involved in the pathophysiology of depression remain poorly understood.

Chapter 1: An overview of Major Depressive Disorder (MDD) from epidemiological and clinical perspectives with a summary of the current knowledge of the underlying biology is provided. A review of the major pathophysiological hypotheses of MDD highlights a need for a more comprehensive approach that allows studying complex molecular interactions involved in depression.

Chapter 2: Transcriptome signature of depression was examined using the measure of replication at individual gene level across different tissues and cell types in both brain and periphery. Fifty-seven replicated genes were reported as differentially expressed in the brain and 21 in peripheral tissues. In-silico functional characterisation of these genes was provided, implicating shared pathways in a comorbid phenotype of depression and cardiovascular disease.

Chapter 3: The molecular basis of MDD using co-expression network analysis was investigated. The Weighed Gene Co-expression Network Analysis (WGCNA) allowed for studying complex interactions between individual genes influencing biological pathways in MDD. Utilising the Sydney Memory and Aging Study (sMAS) and the Older Australian Twin Study (OATS) as discovery and replication cohorts respectively, it was found that the eigengenes of four clusters containing over 3,000 highly co-regulated genes are involved in 13 immune- and pathogen-related pathways and associated with recurrent MDD. However, the findings were not replicated on an independent cohort at the network

level.

Chapter 4: Using a machine learning (ML) approach, a predictive model was built to identify the genome-wide gene expression markers of recurrent MDD. Fuzzy Forests (FF) is a novel ML algorithm, which works in conjunction with WGCNA and was designed to reduce the bias seen in feature selection caused by the presence of correlated transcripts in transcriptome data. FF correctly classified 63% of recurrently depressed individuals in test data using the single top predictive feature (*TFRC,* encodes for transferrin receptor). This suggests that *TFRC* can represent a putative marker for recurrent MDD.

Chapter 5: Following the findings on immune-related pathways being associated with recurrent MDD in the elderly (Chapter 3), the role of these pathways in recurrent MDD was examined at individual gene levels in an independent cohort (OATS). To target the immune pathways, all known genes (KEGG) involved in these 13 pathways were selected and a differential expression analysis was conducted on 1,302 candidates between individuals with recurrent MDD and those without. We found that *CD14* was significantly downregulated in recurrent MDD (FDR < 5%). Considering the key role of *CD14* for facilitating the innate immune response, we suggest that *CD14* can potentially serve as a peripheral marker of immune dysregulation in recurrent MDD.

Chapter 6: A discussion on obtained findings is provided and future directions are outlined with a particular focus on how co-expression network and machine learning approaches that can enhance translation of molecular findings into clinical translation.

Statement

I, Liliana Ciobanu, certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Scholarship.

Liliana Ciobanu, 19/06/2019

Acknowledgements

Firstly, I would like to express my sincere gratitude to my primary supervisor Prof. Bernhard T. Baune for the continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me during my research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study.

Besides my primary supervisor, I would like to thank my co-supervisor Dr. Catherine Toben, for her insightful comments and encouragement, but also for the hard questions which inspired me to widen my research perspectives.

My sincere thanks also go to Dr. David Stacey and Associate Prof. Sarah Cohen-Woods, who had an enormous role in establishing this project. Without their support it would not have been possible to conduct this research.

I would also like to thank the project collaborators Prof. Perminder S. Sachdev, Prof. Julian N. Trollor, Dr. Karen A. Mather, Dr. Anbupalam Thalamutu, and Dr. Simone Reppermund for providing the data from the Sydney Memory and Aging Study (sMAS) and Older Australian Twins Study (OATS) and for the amazing sMAS and OATS team support and hospitality during my 2-week visit of the Centre for Healthy Brain Ageing (CHEBA), UNSW, Sydney.

I would like to acknowledge Adelaide Medical School Discipline of Psychiatry research team for stimulating scientific discussions, suggestions and support.

Last but not the least; I would like to thank my family: Andrei Ciobanu and our three beautiful children – Alexei, Iurie and Ivan. Without your support and encouragement all this would never have happened.

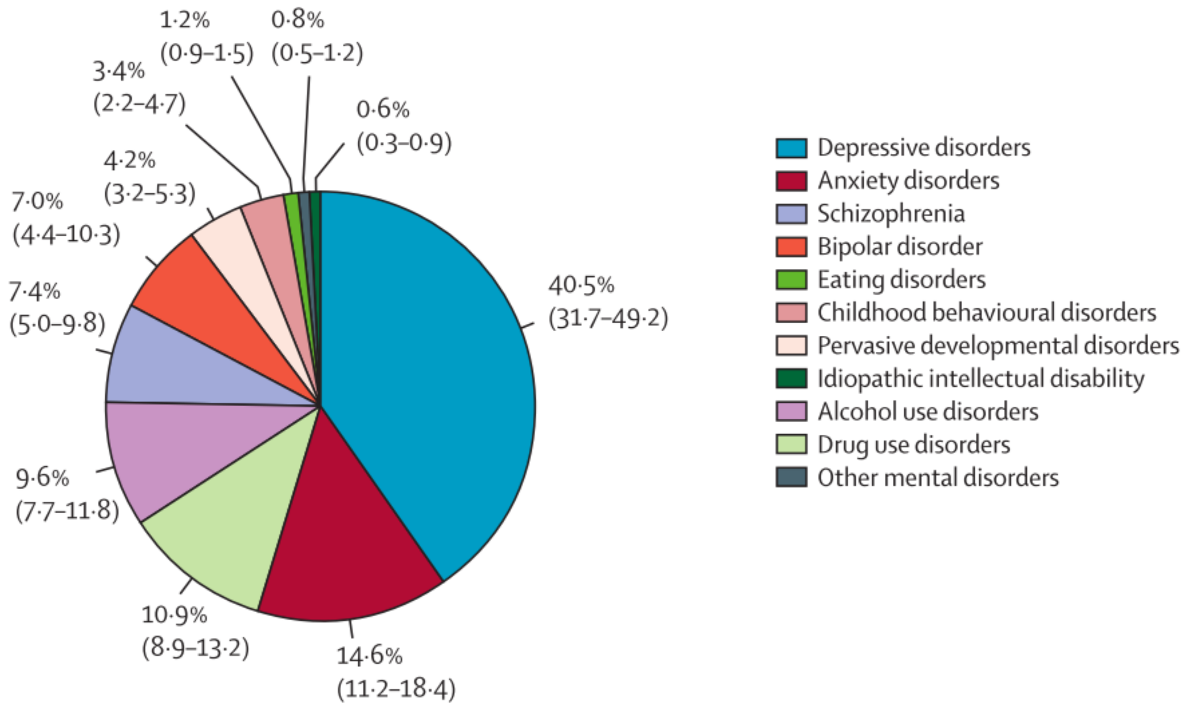
CHAPTER 1

General introduction

EPIDEMIOLOGY OF MAJOR DEPRESSIVE DISORDER (MDD)

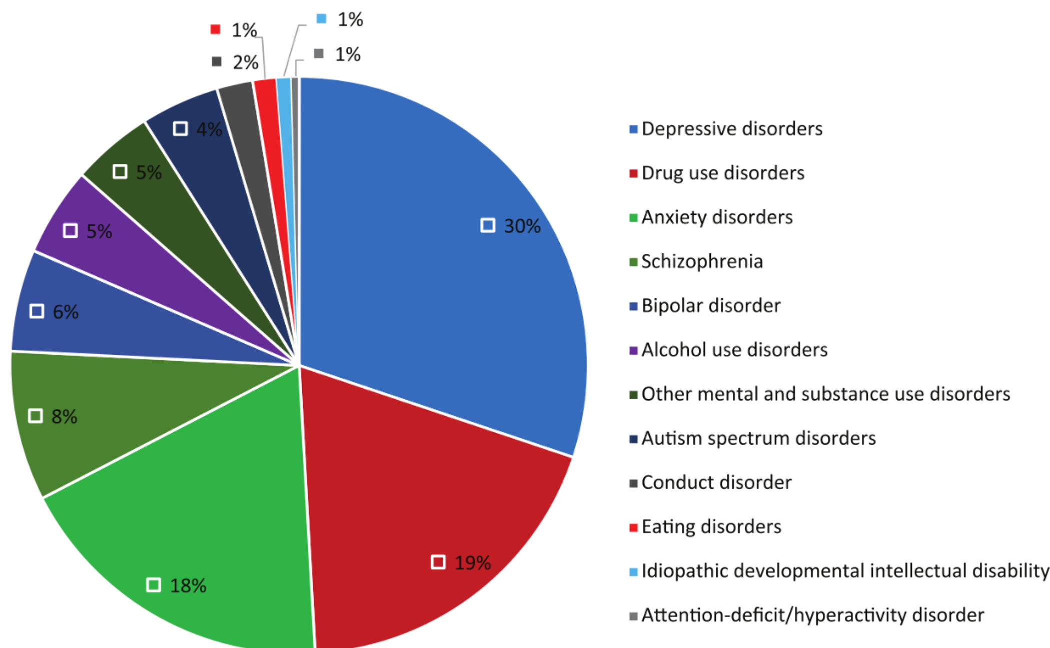
Major Depressive Disorder is a leading cause of disability worldwide, and is one of the major contributors to the overall global burden of disease (Whiteford et al., 2015), including Australia (Ciobanu et al., 2018a) (Appendix A). Globally, more than 300 million people of all ages suffer from depression (WHO, 2017). It has been estimated that depressive disorders (including MDD and dysthymia) are the top contributors to the disease burden attributable to mental and substance use disorders globally (Figure 1.1) as well as in Australia (Figure 1.2) explaining 40.5% and 30% of Disability-Adjusted Life Years, DALYs, respectively (Hay et al., 2017).

Figure 1.1. Proportions of DALYs explained by mental and substance use disorders globally in 2010. Disability-adjusted life years (DALYs) is a sum of years lived with disability (YLDs) and years of life lost (YLLs).



Source: 2010 Global Burden of Disease Study. Seattle, Washington University Institute for Health Metrics and Evaluation (Whiteford et al., 2013)

Figure 1.2. Proportions of DALYs for mental and substance use disorders in Australia in 2015. Disability-adjusted life years (DALYs) is a sum of years lived with disability (YLDs) and years of life lost (YLLs)



Source: 2015 Global Burden of Disease Study. Seattle, Washington University Institute for Health Metrics and Evaluation, 2015 (Ciobanu et al., 2018a)

MDD is also a recognised risk factor for other health outcomes. MDD is an important contributor of burden allocated to ischemic heart disease (Ferrari et al., 2013). Furthermore, individuals suffering from MDD have a shorter life expectancy than those without MDD, in part due to suicidal ideation (Cassano and Fava, 2002).

Considering the burden of this devastating disorder on society, there have been substantial financial investments into improved medical services for affected individuals (Greenberg et al., 2015), which has led to fourfold return in better health and ability to work, according to a WHO-led study (Chisholm et al., 2016). Despite these efforts the global burden of depressive disorders remains largely unchanged between 1990 and 2016 (GBD, 2017).

Clinical Manifestation of Major Depressive Disorder (MDD)

Major depression is a psychiatric mood disorder that goes beyond the normal human experiences of sadness. It encompasses a broad range of symptoms such as feeling worthless, having thoughts of suicide, losing interest in most or all activities, experiencing a significant change (decrease or increase) in appetite or sleep patterns, and having difficulty concentrating. Long-term symptoms can cause clinically significant distress to the individual or lead to impairment in social, occupational, or other important areas of functioning. Clinical diagnosis of MDD uses a symptom-based approach as defined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-V). According to DSM-V, a diagnosis of major depressive disorder is made when the following symptoms are present almost every day for at least two weeks (Box 1).

Box 1

DSM-5 diagnostic criteria for major depression

- A.
  Five (or more) of the following symptoms present during the same 2-week period and which represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) anhedonia
- B.
  Symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning
- C.
  Episode is not attributable to the physiological effects of substance abuse or another medical condition
    1. Depressed mood most of the day (e.g., feels sad, empty, hopeless)
    2. Markedly diminished interest or pleasure in almost all activities nearly every day
    3. Significant appetite changes or significant weight loss or gain
    4. Insomnia or hypersomnia nearly every day
    5. Psychomotor agitation or retardation
    6. Fatigue or loss of energy
    7. Feelings of worthlessness or excessive guilt
    8. Diminished ability to think or concentrate or indecisiveness

The value of disease diagnosis is limited by its reliability, i.e. the agreement between clinicians on making the same diagnosis in the same patient. Reliability is typically evaluated with the kappa coefficient, which ranges from 0 to 1 (from chance to perfect agreement). Benchmarks have been proposed with values above 0.6 considered to be good or very good, between 0.4 and 0.6 moderate, 0.2 to 0.4 fair, and below 0.2 poor (Altman, 2006, Landis and Koch, 1977). The DSM-V field trials, which are designed to ensure that diagnoses were carried out in a way that is representative of psychiatric practice and with an appropriate level of training in the use of DSM-V, have yielded a kappa of 0.28 (95% CI 0.20- 0.35) based on separate interviews by physicians (Darrel A. Regier et al., 2013). Given a low validity of DSM-V diagnosis, i.e. poor agreement between clinicians on symptom-based diagnosis (Uher et al., 2014), greater promise can be expected with biologically based diagnostic markers of MDD that are objective in both methodology and interpretation (Smith et al., 2013).

Limited efficacy of antidepressant treatment

There has been a long debate in the medical community about the effectiveness of currently available antidepressants, centering around whether the observed results in patients can be attributed to the placebo effect (Kirsch, 2014). The largest meta-analysis of 522 trials of 21 antidepressants in 116,477 participants compared efficacy and acceptability of ADs for treatment of adults with moderate to severe MDD has shown that all ADs were more efficacious than placebo (Cipriani et al., 2018). However, rates of total remission following antidepressant treatment are estimated to be only 50.4% (Papakostas, 2010). Because existing first-line antidepressants - classically thought to modulate monoamine neurotransmission - are often insufficient for many patients, there is a greater requirement for improvement in pharmacological antidepressant treatments. The next generation of mechanistically novel therapeutic strategies needs to be more

effective, rapid acting and better tolerated than currently available medications.

Given the increased health and economic burden of depression on society, there is a pressing need for alternative lines of intervention. These can be achieved via developing and implementing improved clinical diagnostic tools and novel therapeutic strategies and treatments. Development of novel clinical diagnostic tools and therapies is paramount in moving towards improved clinical response. A better understanding of the underlying biological pathophysiological mechanisms of depression are required to improve treatment response and predict response outcome. To develop unbiased biology-based diagnostic tools and improved pharmacological treatments, there is a need to better understand that molecular basis of depression. Although decades of experimental research have provided several major biological hypotheses of depression, a comprehensive understanding of the biological correlates of depression remain to be determined. The major hypotheses of pathophysiology of MDD will be covered in the next section.

MAJOR HYPOTHESES OF PATHOPHYSIOLOGY OF MDD

Monoamine deficiency

One of the early biological hypotheses ensued after observations in patients being treated for hypertension with reserpine. Acting as an antagonist reserpine blocks the vesicular monoamine transporter (VMAT) and thereby reduces monoamine levels within the brain. As a result patients were experiencing comorbid depression (Freis, 1954). Following this observation, the hypothesis that a deficiency or imbalance in the monoamine system of the brain is an underlying biological basis for depression was proposed. Monoamines refer to the particular neurotransmitters dopamine, norepinephrine and/or serotonin. On the basis of this hypothesis, various antidepressants have been designed to increase the

levels of monoamines within the synaptic cleft either via inhibition of monoamine degradation or by the blockade of their reuptake (Slattery et al., 2004). Currently marketed antidepressants have the monoamine hypothesis as their theoretical basis. The two original antidepressants were the monoamine oxidase inhibitor (MAOI) Iproniazid, and the tricyclic antidepressant (TCA) Imipramine. The determination of their mode of action led to the catecholamine hypothesis of depression being developed in the mid-1960s (Schildkraud, 1965). Monoamine elevation therapies have for the large part proved successful and remain the most widely prescribed pharmaceuticals (e.g. TCAs, MAOIs and uptake inhibitors); however, two caveats remain as all monoaminergic antidepressants have a delayed onset of action of several weeks and therapeutic unresponsiveness is indicated in roughly 30% of depressed patients (Doris et al., 1999, Machado-Vieira et al., 2008). Given the limited success of monoamine interventions and the increasing burden of the disorder on society, novel non-monoamine interventions have emerged.

HPA axis hyperactivity

The hypothalamic-pituitary-adrenal (HPA) axis is the fundamental neuroendocrine system that controls reactions to stress. This axis consists of stimulating forward and feedback inhibition loops involving the brain, pituitary, and adrenal glands, which regulate glucocorticoid production. Cortisol released from the adrenal glands, binds within the brain with high affinity to mineralocorticoid receptors (MRs) and with lower affinity to glucocorticoid receptors (GRs). The hyperactivation of the HPA axis reflecting a dysregulation of MR and/or GR is one of the most consistent findings in neurobiology of depression, but the mechanisms underlying this abnormality are still unclear. This increased activity of the HPA axis is thought to be related, at least in part, to reduced feedback inhibition by endogenous glucocorticoids (Pariante and Lightman, 2008).

Altered neural plasticity and neurogenesis

Neural plasticity is a fundamental mechanism of neuronal adaptation. It is becoming increasingly clear that altered neurogenesis and neural plasticity induced by stress and other negative stimuli play a significant role in the onset and development of depression (Duman et al., 1999). According to this theory, neural circuits and connections undergo lifelong modifications and reorganizations in response to external or internal environmental stimuli. Adult neurogenesis involves precursors of cell proliferation, migration and differentiation mainly occurring in the dentate gyrus of the hippocampus (Eriksson et al., 1998). Neurotoxic agents such as chronic stress, excessive concentrations of glutamate, biogenic amines and glucocorticoids may affect the morphology of some neural cells such as hippocampal CA3 pyramidal neurons and pyramidal cells of prefrontal cortex (Serafini, 2012). Reduced hippocampal volume is one of the most common findings in depressed individuals and longer duration of depressive episodes is known to be closely related to modifications in hippocampal volume (Lorenzetti et al., 2009). However, there is no scientific consensus to confirm the direction of causality. Most of the studies suggest that depression and dysfunction of neural plasticity act on and influence each other (Liu et al., 2017).

Dysregulation of glutamatergic system

Glutamate is the most abundant excitatory neurotransmitter in the central nervous system. It is found in substantially higher concentrations than monoamines and in more than 80% of neurons, highlighting its role as a major excitatory synaptic neurotransmitter (Mathew et al., 2005). Given that glutamate is so widely distributed in the brain, strict regulation is necessary to prevent undue excitotoxicity. The delicate balance of glutamate with the major inhibitory neurotransmitter γ-aminobutyric acid (GABA) is essential for all physiological homeostasis in the CNS (Schoepp, 2001). Early findings within the 1990s

showed that N-methyl-D-aspartate receptor (NMDA-R) antagonists possess antidepressant-like action (Trullas and Skolnick, 1990) leading to the 'glutamate hypothesis of depression'. Later studies have found that MDD pathophysiology is associated with dysfunction of the predominant glutamatergic system, malfunction in the mechanisms regulating clearance and metabolism of glutamate, and morphological maladaptive changes in a number of limbic/cortical areas in the brain mediating cognitive-emotional behaviours (Sanacora et al., 2012). Glutamate itself serves as a metabolic precursor for the neurotransmitter GABA, via the action of the enzyme glutamate decarboxylase. The role of GABA specifically in MDD is briefly discussed below.

Reduced GABAergic activity

GABA (γ-aminobutyric acid) is the chief inhibitory neurotransmitter in the mammalian central nervous system. Its principal role is reducing neuronal excitability throughout the nervous system. The GABAergic deficit hypothesis of depression posits that reduced GABA concentrations in the brain, impaired function of GABAergic interneurons, altered expression and function of $GABA_A$ receptors, and changes in GABAergic transmission dictated by altered chloride homeostasis can contribute to the aetiology of MDD (Luscher and Fuchs, 2015).

Dysregulation of melatonergic system

Melatonin (*N*-acetyl-5-methoxytryptamine) is a neurohormone that is prominently, albeit not exclusively, synthesised in the pineal gland and secreted in a phasic manner (its circulating level varies in a daily cycle). Melatonin exerts its actions through membrane MT1/MT2 melatonin receptors, which belong to the super family of G-protein-coupled receptors and are primarily expressed in the CNS (De Berardis et al., 2013). Circadian rhythms, regulated by the melatonergic system, have long been considered to be

disrupted in MDD, bringing on depressive behaviours and symptoms, disrupted sleep and poor regulation of neuroendocrine mediators such as cortisol, norepinephrine (NE) and serotonin (McClung, 2007). However, the complex relationships between the circadian system and the development of depressive symptoms are far from being elucidated (Courtet and Olie, 2012). Stimulation of melatonergic (MT_1 / MT_2) receptors by melatonergic antidepressants, such as agomelatine, purported to resynchronize circadian rhythms, was found moderately more effective than placebo with similar efficacy and fewer side effects to standard antidepressants in the treatment of depressed patients (Cardinali et al., 2013, Taylor et al., 2014).

Inflammation in depression

Finally, pathophysiology of depression was associated with the immune system and inflammation (Maes, 1999). Extensive findings support the role of chronic low grade inflammation in depression. MDD patients exhibit all of the cardinal features of an inflammatory response, including increased expression of pro-inflammatory cytokines and their receptors (McAfoose and Baune, 2009, Mills et al., 2013) and increased levels of acute-phase reactants (Wium-Andersen et al., 2013, Kohler-Forsberg et al., 2017), chemokines (Eyre et al., 2016, Singhal and Baune, 2018) and soluble adhesion molecules in peripheral blood and cerebrospinal fluid (CSF) (Lespérance et al., 2004, Dimopoulos et al., 2006). Inflammation in depressed patients has been linked to altered gut microbiota dysbiosis (dysregulation of gut-brain-axis) (Clapp et al., 2017, Schachter et al., 2018), childhood trauma (Cattaneo et al., 2015), and stress-related epigenetic regulations (Wang et al., 2018). Inflammation has also been observed in depressed patients suffering from obesity (Ambrosio et al., 2018), cardiovascular disease (Halaris, 2017), cancer (Li et al., 2017, Weber and O'Brien, 2017), and asthma (Jiang et al., 2014), suggesting inflammation to play an important role in comorbid depression. Accumulating

evidence suggests that anti-inflammatory treatments are associated with anti-depressant properties. Although the therapeutic effect of anti-inflammatory drugs has been observed in several clinical trials (Köhler et al., 2014, Köhler et al., 2016), it has been mainly explored in acute disease stages, which indicates its potentially limited application in clinical settings (Baune, 2018).

Although the empirical evidence for the role of the immune system in depression is increasing, our understanding of the immunology underlying inflammation in depression is limited. For instance, not all patients with increased inflammation develop MDD nor do all people with MDD show prominent immune activation (Steptoe et al., 2003). Therefore, it is becoming increasingly clear that dysregulated inflammatory responses are not necessary or sufficient for the development of depression, as inflammatory markers appear to be increased only in a subset of patients (Rosenblat et al., 2014). Furthermore, while both innate and adaptive immune systems seem to be involved in depression, the interplay between the two remains unclear.

WHY STUDY THE TRANSCRIPTOME IN DEPRESSION?

Surmounting evidence suggests that depression is a multifaceted disorder with both genetic and environmental factors contributing to the onset and progression of the disorder. Despite substantial heritability of depression estimated at 31% to 42% (Sullivan et al., 2000), identification of the genetic underpinnings of depression has been challenging. An intensive search for genetic factors of depression using a candidate gene approach pointed towards more than 200 genetic loci, mainly genes involved in neurotransmission and the hypothalamic–pituitary–adrenal axis (HPA), however, only a few of these findings have been successfully replicated (Rivera and McGuffin, 2015). Genome-wide association studies (GWAS), after several unsuccessful attempts (Flint and Kendler, 2014) recently revealed 18 novel loci associated with depression at genome-wide level. Two loci were found associated with severe depression in Han Chinese women (CONVERGE, 2015), and 15 loci were identified through 23andMe using self-report data of severely depressed individuals (Hyde et al., 2016), and one recently identified locus was found to be associated with late-onset depression (Power et al., 2017). Although each of the three GWAS studies validated their findings within replication studies, there was no overlap in genetic variants across these studies. The discrepancies between the findings reflects the highly heterogeneous nature of depression (Levinson et al., 2014).

Studying global gene expression is a relatively novel and promising approach to uncover the pathophysiology of depression as well as to possibly provide useful clinical information for predicting treatment response and identification of appropriate treatment options. Quantifying the abundance of mRNA molecules in a single cell or from within a population of cells provides essential information on the biological activity and functions of genes. Studying gene expression in depression can be viewed as being complementary to a gene discovery approach aimed at understanding the dynamic

molecular changes in depression. Given that the level and patterns of gene expression are influenced by both genetic and environmental factors (Wright and Sullivan, 2014), - such as age (van den Akker et al., 2014), sex (Jansen et al., 2014), smoking status (Charlesworth et al., 2010) and well-being (Fredrickson et al., 2013), - association between gene expression and depression may reflect an interactive effect of both. Within a clinical research context, the identification of altered gene expression patterns in depression is of critical importance for (1) a better understanding of molecular underpinnings of depression, (2) establishing biological clinical markers of depression, (3) increasing the evidence-base for the development of novel antidepressants, and (4) identifying biomarkers for predicting treatment outcome, all of which are urgently needed for a better diagnosis and for more personalised treatments of affected individuals (Ferrari et al., 2013).

Rapidly advancing technologies, such as microarrays and RNA-sequencing that allow for transcriptome coverage have become powerful tools to quantify levels of gene expression in various tissues relevant for the pathophysiology of depression. In the next chapter, the gene expression signature of depression is explored using the measure of replication at the individual gene level. Investigation includes extraction of a comprehensive list of those genes which were found to be dysregulated in depression in both brain and periphery across the lifespan and then determination of the molecular pathways these genes are involved in, and to what extent brain and peripheral tissues/cells findings overlap.

Statement of Authorship

Title of Paper	Genome-wide differential gene expression in brain and peripheral tissues in MDD across the life span
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	CIOBANU, L. G., SACHDEV, P. S., TROLLOR, J. N., REPPERMUND, S., THALAMUTHU, A., MATHER, K. A., COHEN-WOODS, S. & BAUNE, B. T. 2016. Differential gene expression in brain and peripheral tissues in depression across the life span: A review of replicated findings. *Neuroscience & Biobehavioral Reviews*, 71, 281-293

Principal Author

Name of Principal Author (Candidate)	CIOBANU, L. G.		
Contribution to the Paper	Conceived of the presented research concept and design, performed a literature review, made interpretations of the results and prepared the first draft of the manuscript. Made manuscript submission, prepared a reply letter for reviewers and finalised the manuscript based on reviewers' feedback.		
Overall percentage (%)	85%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	20. 05. 2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	SACHDEV, P. S.		
Contribution to the Paper	• Revised the draft paper		
Signature		Date	11 June 2019
Name of Co-Author	TROLLOR, J. N.		
Contribution to the Paper	• Revised the draft paper		
Signature		Date	11/06/2019
Name of Co-Author	REPPERMUND, S.		

Contribution to the Paper	• Revised the draft paper		
Signature		Date	11/06/2019
Name of Co-Author	THALAMUTHU, A.		
Contribution to the Paper	• Revised the draft paper		
Signature		Date	12/06/2019
Name of Co-Author	MATHER, K. A.		
Contribution to the Paper	• Revised the draft paper		
Signature		Date	6/6/19
Name of Co-Author	COHEN-WOODS, S.		
Contribution to the Paper	• Revised the draft paper		
Signature		Date	28/05/2019
Name of Co-Author	BAUNE, B. T.		
Contribution to the Paper	• Supervised the project • Provided critical feedback to study design and methodology • Provided critical feedback to the first draft of the manuscript • Discussed the results and contributed to the final manuscript • Final decision on manuscript submission		
Signature		Date	16.06. 2019

CHAPTER 2

Genome-wide differential gene expression in brain and peripheral tissues in MDD across the life span

Neuroscience and Biobehavioral Reviews 71 (2016) 281–293

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Differential gene expression in brain and peripheral tissues in depression across the life span: A review of replicated findings

Liliana G. Ciobanu[a], Perminder S. Sachdev[b], Julian N. Trollor[b], Simone Reppermund[b,c], Anbupalam Thalamuthu[b], Karen A. Mather[b], Sarah Cohen-Woods[d], Bernhard T. Baune[a,*]

[a] *Discipline of Psychiatry, School of Medicine, University of Adelaide, South Australia, Australia*
[b] *Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, University of New South Wales, New South Wales, Australia*
[c] *Department of Developmental Disability Neuropsychiatry, School of Psychiatry, University of New South Wales, New South Wales, Australia*
[d] *School of Psychology, Faculty of Social and Behavioural Sciences, Flinders University, South Australia, Australia*

A R T I C L E   I N F O

Article history:
Received 10 May 2016
Received in revised form 25 July 2016
Accepted 16 August 2016
Available online 24 August 2016

Keywords:
Depression
Major depressive disorder
Transcriptome
Genome-wide differential gene expression

A B S T R A C T

There is a growing body of research investigating the gene expression signature of depression at the genome-wide level, with potential for the discovery of novel pathophysiological mechanisms of depression. However, heterogeneity of depression, dynamic nature of gene expression patterns and various sources of noise have resulted in inconsistent findings. We systematically review the current state of transcriptome profiling of depression in the brain and peripheral tissues with a particular focus on replicated findings at the single gene level. By examining 16 brain regions and 5 cell types from the periphery, we identified 57 replicated differentially expressed genes in the brain and 21 in peripheral tissues. Functional overlap between brain and periphery strongly implicates shared pathways in a comorbid phenotype of depression and cardiovascular disease. The findings highlight dermal fibroblasts as a promising experimental model for depression biomarker research, provide partial support for all major theories of depression and suggest a novel candidate gene, *PXMP2*, which plays a critical role in lipid and reactive oxygen species metabolism.

© 2016 Elsevier Ltd. All rights reserved.

INTRODUCTION

There is a growing body of research investigating the gene expression signature of depression at the genome-wide level, with potential for the discovery of novel pathophysiological mechanisms of depression. However, heterogeneity of depression, the dynamic nature of gene expression patterns and various sources of noise have

resulted in inconsistent findings. We systematically review the current state of transcriptome profiling of depression in the brain and peripheral tissues with a particular focus on replicated findings at the single gene level. By examining 16 brain regions and 5 cell types from the periphery, we identified 57 replicated differentially expressed genes in the brain and 21 in peripheral tissues. Functional overlap between brain and periphery strongly implicates shared pathways in a comorbid phenotype of depression and cardiovascular disease. The findings highlight dermal fibroblasts as a promising experimental model for depression biomarker research, provide partial support for all major theories of depression and suggest a novel candidate gene, *PXMP2,* which plays a critical role in lipid and reactive oxygen species metabolism.

Genome-wide gene expression in depression

The application of high-throughput gene expression analyses has gained momentum in the study of molecular signatures of diseases. Microarray and next generation sequencing (NGS) technologies, which permit profiling the expression of many thousands of genes simultaneously, have been applied with success in many areas, including cancer research. Genomic and transcriptome alterations have enabled molecular classifications of cancer and revealed novel biomarkers for diagnosis, prognosis, and predicting response to therapies (Roychowdhury and Chinnaiyan, 2016) and inspired many other fields of medical research to utilize newly developed techniques. In recent decades, the field of psychiatry has adopted these techniques, aiming to elucidate molecular mechanisms, identify biomarkers and provide better treatment for depression, the leading cause of disability affecting more than 350 million people worldwide (WHO, 2015).

The problem of non-replication of gene expression findings

A growing amount of genome-wide gene expression data has been analysed using differential expression analysis, the most widely applied statistical method. However,

numerous limitations of biological and technical nature, including large biological variations, small sample sizes, data collection details, clinical heterogeneity, comorbidities, differences in microarray platforms, data quality assessment, statistical algorithms used and covariates accounted for, and many others, have resulted in inconsistent results, questioning their validity. Biological findings need to be confirmed by several studies using the same method in order to be accepted. While the lack of replication is a major concern for transcriptome studies in depression, the systematic collection of replicated findings have never been performed. We address this gap by exploring the gene expression signatures of depression derived from both brain and peripheral tissues using replication as the yardstick of reliability.

The choice of tissue for depression in gene expression research

Depression includes dysfunction at multiple biological levels, from genes (Ripke et al., 2013) to brain regions (Gong and He, 2015) and blood circulating throughout the body (Lopresti et al., 2014). The choice of tissue, therefore, is of particular importance in gene expression research. Studies performed on post-mortem brains have substantially advanced our understanding of the pathophysiological mechanisms of depression. Gene expression signatures derived from various brain regions collectively point towards various molecular processes involving inflammatory, cell survival, apoptotic, oxidative stress and other pathways (Mehta et al., 2010). However, brain expression findings cannot be used for diagnostic purposes due to limited access to tissue from affected individuals. Extensive research on peripheral biomarkers of depression has revealed that peripheral immune response and growth factors, endocrine factors and metabolic markers also contribute to the pathophysiology of depression (Lin and Tsai, 2016). This is consistent with the close interaction between the brain and peripheral tissues. However, whether gene expression pattern in a peripheral tissue, such as blood, is a reflection of brain activity or a separate peripheral tissue process independent of the brain, remains

to be understood. It is therefore necessary not only to examine peripheral gene expression but also to compare the brain and periphery gene expression findings to address some of these questions in depression research.

The main challenge is to compile the numerous transcriptome profiles derived from different brain areas or/and peripheral cell types into one coherent analysis in an attempt to explain the mechanisms of depression. In this review, we compare transcriptomes obtained from multiple cell types in order to identify replicated findings. It can be argued that if any particular gene, in the face of various biological and technical limitations, was differentially expressed in depression compared to healthy controls across several cell/tissue types or brain areas, this gene has an increased likelihood of being truly involved in the pathophysiology of depression. We explore the gene expression signature of depression using replicability at the single gene level as a method of maximising true associations.

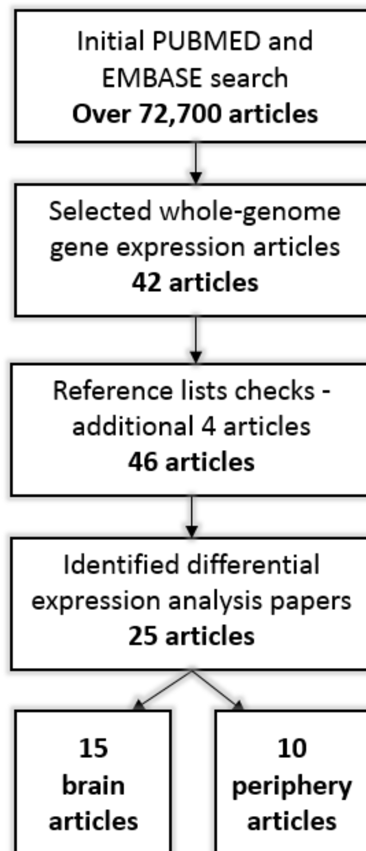
METHODS

Article selection process

Using PubMed and EMBASE databases, we screened for all gene expression studies in depression in humans published in peer-reviewed journals using various permutations of the following search terms: “transcriptome”, “gene expression”, “depression”, “MDD”, “Major Depressive Disorder”. This preliminary literature search resulted in over 72,700 articles. In the second step, based on the titles and information provided in abstracts, we selected 42 articles with genome-wide expression data. The inclusion criteria for the review were: (1) phenotype of depression, (2) original genome-wide gene expression data, and (3) differential expression analysis between depression and controls. We did not restrict our selection by sample characteristics, like age, gender, and ethnicity. Using

the reference lists checks, we obtained additional 4 articles. Finally, we identified studies that utilized differential expression analysis between depression and controls. This systematic search strategy allowed us to select 15 articles on the brain (not limited to specific brain areas) and 10 on peripheral tissues (Figure 2.1).

Figure 2.1. Selection process flowchart

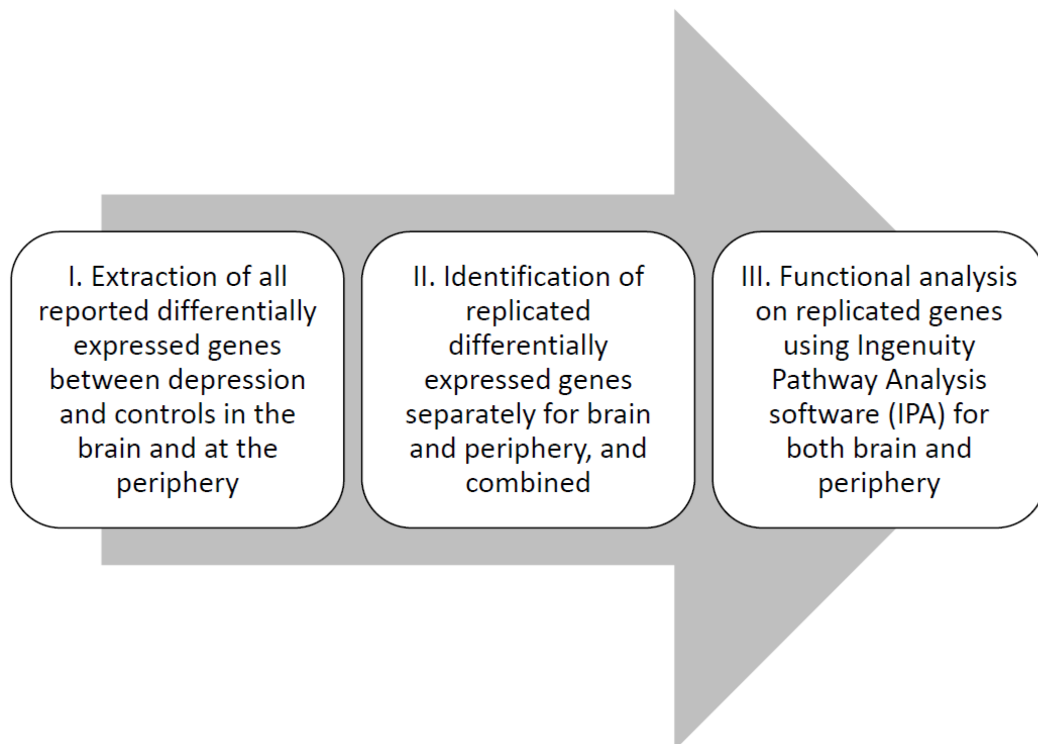


Measure of replication as a review method

Given that calculating the percentage of overlapping genes (POGs) to evaluate the replicability of the results across different studies is not a valid approach in a situation with unequal lengths of differentially expressed gene (DEG) lists (Zhang et al., 2009), we used replication of single differentially expressed gene across the studies both separately and together for brain and peripheral tissues studies. First, we manually extracted all DEGs between depression and healthy controls reported by each individual study and,

using the multi-symbol checker (HUGO Gene Nomenclature Committee, HGNC), checked if these genes were known by another name (synonyms, previous names). Second, we identified all replicated DEGs for brain regions and peripheral tissues. Third, we performed functional annotation of replicated DEGs using Top Diseases and Functions Network Analysis (Ingenuity Pathway Analysis, IPA) (Figure 2).

Figure 2.2. Workflow chart



RESULTS

Mapping the transcriptome signature of depression in the brain

The complexity of brain function, the heterogeneous phenotype of depression and the inevitable limitations of post-mortem studies together create a serious problem of integration of multiple single gene findings into a systematic network level translatable into observable behaviours of depression. In this review, we attempt to compile multiple differential gene expression findings in depression and controls across different brain

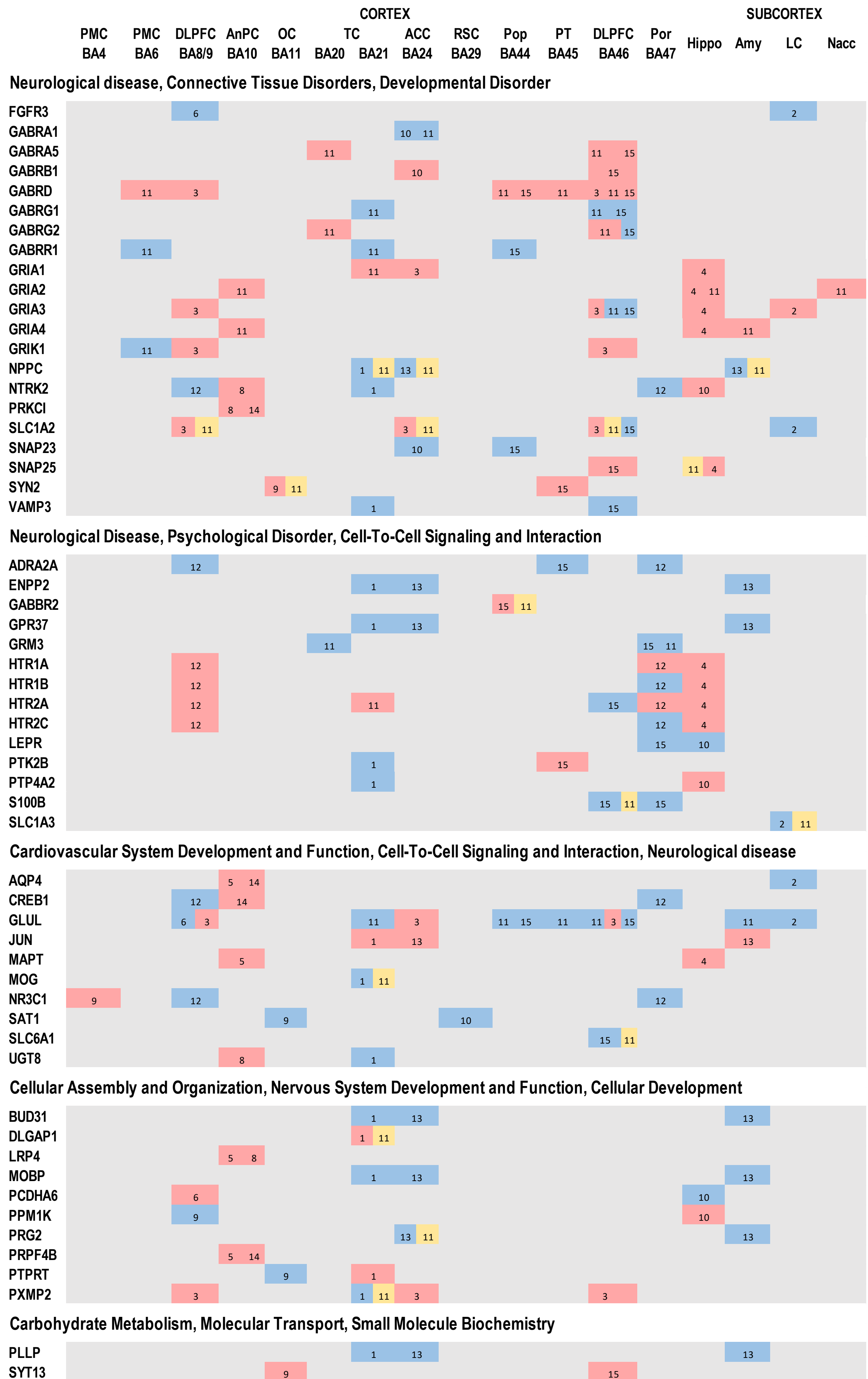
regions. Transcriptomes from 12 cortical and 4 subcortical areas were derived using microarray and RNA-seq technologies. Over 500 genes were reported to be dysregulated in depression across the brain in 15 studies selected for this review. However, small sample sizes (ranging from 9 to 21 for MDD cases across 15 studies) in combination with whole-genome statistics make these findings sensitive to false positive results. Tracking genes replicated across multiple studies is one of the ways to increase the validity of the results without losing brain areas specificity.

Replication of gene expression findings in the brain

A thorough examination of 15 brain studies showed that 582 genes were differentially expressed between depression and controls. Without discriminating between the different brain regions there were 57 replicated DEGs (9.8%) (Table 2.1). That is each gene was differentially expressed in any brain area more than once irrespective of the directionality of expression. Only 28 genes out of the combined pool of 582 DEGs (4.8%) were replicated in the same brain area. Among them, 5 genes were dysregulated in opposite directions, which may be partially explained by methodological differences between the studies. Top Diseases and Functions analysis (IPA) of the 57 replicated DEGs suggests that these genes collectively are involved in neurological disease, connective tissue disorders, developmental disorder, psychological disorder, cell-to-cell signalling and interaction, cardiovascular system development and function, cellular assembly and organization, nervous system development and function, cellular development, carbohydrate metabolism, molecular transport, small molecule biochemistry (Table 2.1).

Table 2.1. The 57 replicated differentially genes expressed mapped to the 6 cortical areas in depression (BA 8/9, BA21, BA24, BA44, BA45, BA46). Genes are listed according to the Top Diseases and Functions networks identified by IPA. Blue represents downregulation, pink upregulation, yellow information is not available. Numbers refer to the studies cited (see below)

Transcriptome signature of depression



Abbreviations:

PMC - Primary Motor Cortex, PC – Premotor Cortex, DLPFC – Dorsolateral Prefrontal Cortex, AnPC – Anterior Prefrontal Cortex, OC – Orbital Cortex, TC – Temporal Cortex, ACC – Anterior Cingulate Cortex, RSC – Retrosplenial Cortex, Pop – Pars Opecularis (part of the inferior frontal gyrus and part of Broca's), PT - Pars triangularis (part of the inferior frontal gyrus and part of Broca's area), POr - Pars orbitalis (part of the inferior frontal gyrus), Amy – Amygdala, LC – Locus Coeruleus, Nacc – Nucleus Accumbens, Hippo – Hippocampus

Studies:

[1]Aston et al. (2005), [2]Bernard et al. (2011), [3]Choudary et al. (2005), [4]Duric et al. (2013), [5]Iwamoto et al. (2004), [6]Kang et al. (2007), [7]Kohen et al. (2014a) [8]Malki et al. (2015), [9]Sequeira et al. (2006), [10]Sequeira et al. (2007), [11]Sequeira et al. (2009), [12]Sibille et al. (2004), [13]Sibille et al. (2009), [14]Tochigi et al. (2008), [15]Klempan et al. (2009).

The most replicated differentially expressed genes in the brain

GABAergic and glutamatergic-related genes, which code for two major neurotransmitters in the brain (GABA, glutamate), show the most abundance in the replicated genes map with the highest dysregulation in prefrontal cortical areas, which are well known to be involved in mood regulation and depression (Table 2.1). However, the directionality of expression of GABAergic and glutamatergic-related genes across brain areas was inconsistent.

The most widespread dysregulated gene across the brain is *GLUL, a* glutamatergic-related gene coding for glutamate-ammonia ligase, which has been previously implicated in the pathophysiology of depression. GLUL expression was dysregulated across 6 cortical areas Brodmann Areas 8/9, 21, 24, 44, 45, and 46 (BA8/9, BA21, BA24, BA44, BA45, and BA46) and 2 subcortical areas (amygdala and locus coeruleus). However, the direction of dysregulation across different brain areas was not uniform, suggesting that *GLUL*, known to be involved in many biological functions, may be downregulated in one brain area/cell type and upregulated in another. Also noteworthy, the findings are conflicting for the two dorsolateral prefrontal cortical regions, BA46 and BA8/9. Choudary et al. (2005) found *GLUL* to be upregulated in BA8/9, whereas Kang et al. (2007) showed that *GLUL* is downregulated in BA8/9; similarly, Sequeira et al. (2009) and Klempan et al. (2009) have found *GLUL* to be downregulated in BA46, while Choudary et al. (2005)

found it upregulated. This could possibly be explained by the differences in phenotype: MDD with unspecified cause of death (Choudary et al., 2005) vs. SMD – suicides with major depression (Sequeira et al., 2009, Klempan et al., 2009).

The glutamate transporter gene *SLC1A2*, which has been previously shown to be dysregulated in depression, shows consistent dysregulation of gene expression across three brain areas: the dorsolateral prefrontal cortices (BA8/9, BA46) and anterior cingulate cortex (BA24). Although dysregulation in BA8/9 and BA24 has been replicated twice and in BA46 three times, the direction of dysregulation is inconsistent. Choudary et al. (2005) found *SLC1A2* to be upregulated in BA8/9, BA24 and BA46, whereas Klempan et al. (2009) found it to be downregulated in BA46. The directionality of dysregulation of *SLC1A2* in BA8/9, BA24 and BA46 identified by Sequeira et al. (2009) was not reported.

Another gene showing dysregulation across multiple brain areas is the *GABRD* - GABAergic gene, coding for gamma-aminobutyric acid type A receptor delta subunit, which was previously shown to play a role in depression (Feng et al., 2010). *GABRD* expression was consistently upregulated across 5 brain areas (BA6, BA8/9, BA44, BA45, and BA46) irrespective of phenotype (MDD vs. SMD) or sex (males only in Sequeira et al. (2009), Klempan et al. (2009), and both sexes in Choudary et al. (2005). This strongly suggests that upregulation of *GABRD* in aforementioned brain areas is related to depression for both males and females.

Another group of consistently replicated genes are those from the serotonergic family, *HTR1A, HTR1B, HTR2A, HTR2C,* which were predominantly observed in the dorsolateral prefrontal cortex (BA8/9), pars orbitalis, part of the inferior frontal gyrus (BA47) (Sibille et al., 2004) and hippocampus (Duric et al., 2013). The *HTR2A* gene, in addition to BA8/9, BA47 and hippocampus, was also upregulated in the temporal cortex (BA21) (Sequeira et al., 2009) but downregulated in BA46 (Klempan et al., 2009). While these genes show

replication across different brain regions, none was replicated within the same brain area. Since serotonergic genes are known for their role in suicidal behaviour (Antypa et al., 2013), interpretation of these replications in regards to depression only should be treated with caution as all four studies used the combined phenotype of depression and suicide.

The PXMP2 gene, known to be involved in reactive oxygen species (ROS) metabolism, was replicated by three studies. The level of PXMP2 was found upregulated in the BA8/9, BA24, and BA46 (Choudary et al., 2005), downregulated in BA21 (Aston et al., 2005), and dysregulated in BA21 (Sequeira et al., 2009).

The first RNA-seq study did not replicate microarray findings

An important factor in brain research that is often not taken into account is the cellular heterogeneity of investigated brain areas. From which cells mRNAs are extracted may greatly affect the overall pattern of expression as functional distinct cells may express mRNAs in different amounts. To overcome this limitation, Kohen et al. (2014a) performed the first whole transcriptome sequencing (RNA-seq) of dentate gyrus (DG) granule cells captured by the challenging but powerful laser capture microdissection (LCM). None of 27 genes ($P<0.001$) identified as significantly involved in shaping gene expression differences between depression and controls was replicated by any of the 14 reviewed brain whole-genome microarray studies. This suggests that the gene expression signature of depression is complex and further research on identifying the cell-specific patterns of expression across relevant brain areas is needed.

Transcriptomic signature of depression at the periphery

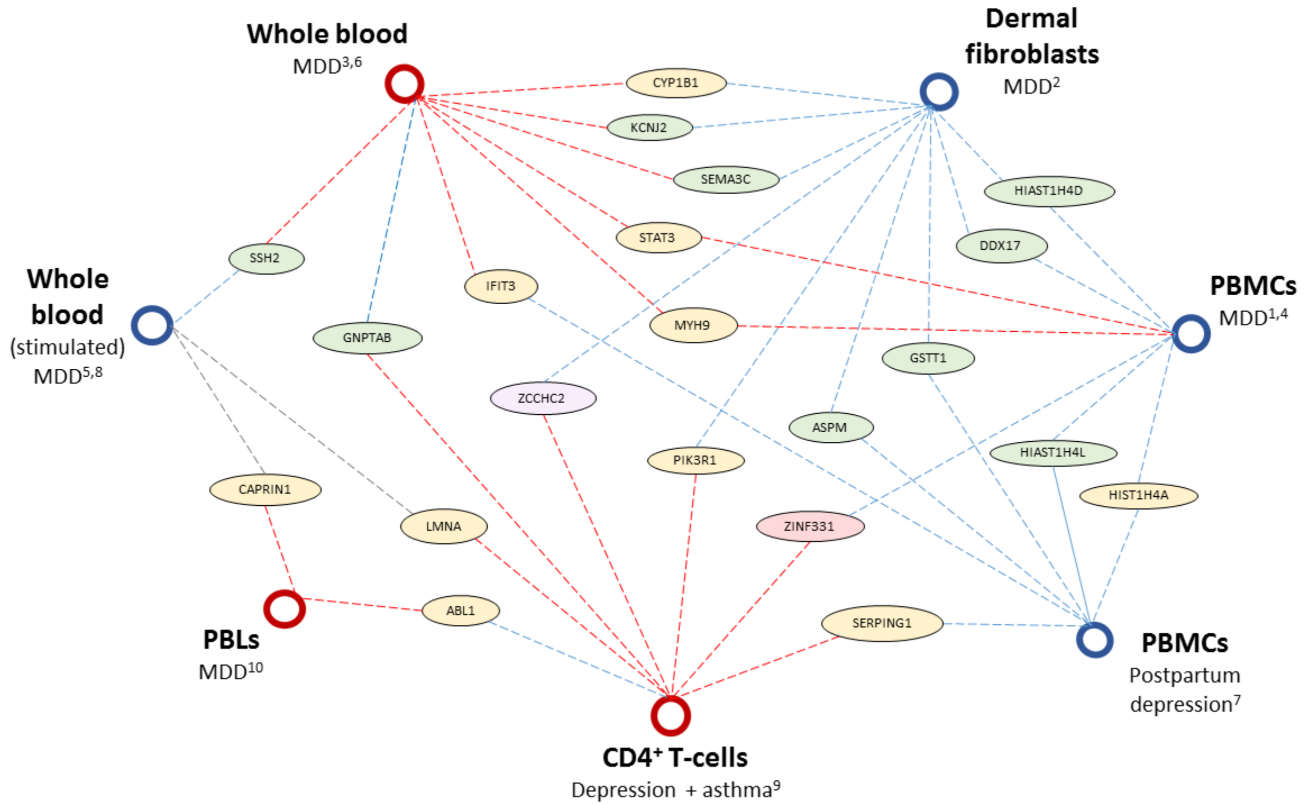
Given that blood and brain cells share 81.9% of the transcriptome (Liew et al., 2006), peripheral blood profiling gives us an opportunity to study some of the aspects of brain functioning in the absence of human neural tissue. Various types of peripheral blood cells

have been investigated in depression and gene expression studies, such as mononuclear cells (PBMCs) (Belzeaux et al., 2012, Liu et al., 2014, Segman et al., 2010), lymphocytes (Yi et al., 2012), CD4+ T cells (Wang et al., 2015), as well as dexamethasone (Menke et al., 2012) and lipopolysaccharides (LPS) (Spijker et al., 2010) stimulated, and unstimulated whole blood.

Out of the combined pool of reported 752 DEGs, only 21 DEGs were replicated in an independent cohort. None of them have been replicated more than once across all reported studies. Moreover, only 8 out of 21 (38.1%) replicated DEGs showed a consistent direction of expression, whereas 11 were expressed in the opposite direction (52.3%), and for 2 genes the information about directionality was not available (9.5%) (Figure 2.3). This inconsistency creates a major obstacle for developing a peripheral blood biomarker panel for depression.

The Top Diseases and Functions analysis (IPA) of the 21 overlapping DEGs identified in depression vs. healthy controls in the periphery highlighted the functional role of these genes in processes related to the cell cycle, cancer, cardiovascular disease, cellular assembly and organisation, cellular function and maintenance, amino acid metabolism, cell and organ morphology, cell death and survival and embryonic development (Figure 2.3).

Figure 2.3. Replicated differentially expressed genes identified in periphery cells in depression. Red dotted lines represent upregulation; blue downregulation; and, grey for directionally of expression not reported.



Abbreviations: MDD – Major Depressive Disorder, PBMCs – peripheral blood mononuclear cells, PBLs – peripheral blood leukocytes

Studies: [1]Belzeaux et al. (2012), [2]Garbett et al. (2015a), [3]Jansen et al. (2016), [4]Liu et al. (2014), [5]Menke et al. (2012), [6]Mostafavi et al. (2014), [7]Segman et al. (2010), [8]Spijker et al. (2010), [9]Wang et al. (2015), [10]Yi et al. (2012).

Top Disease and Functions	Genes
Cell cycle, Cancer, Cardiovascular Disease	ABL1, CAPRIN1, CYP1B1, HIST1H4A, IFIT, LMNA, MYH9, PIK3R1, SERPING1, STAT3
Cellular Assembly and Organization, Cellular Function and Maintenance, Organ Morphology	ASPM, DDZ17, GNPTAB, GSTT1, HIST1H4L, KCNJ2, SEMA3, SSH2
Amino Acid Metabolism, Cell Morphology, Cellular Function and Maintenance	ZCCH2
Cell Death and Survival, Embryonic Development, Organ Morphology	ZNF31

Isolated cells vs. whole blood in biomarker research

Using the measure of replication at the single gene level, we found that overall the results from isolated cell types had a better replication rate than results obtained using whole blood. This may partly be related to differential power, i.e. isolated cells with more specific signals yield high power for the same sample size, in comparison with whole blood, with multiple signals from different cell types. The heterogeneity of depression also reduces statistical power. For instance, depression transcriptomes obtained from PBMCs showed overlap for 6 (Segman et al., 2010) and 7 (Belzeaux et al., 2012) DEGs across all studies with only 9 depression cases. On the other hand, the first deep RNA-seq of whole blood of 463 MDD cases and 459 healthy controls (Mostafavi et al., 2014) showed only 2 overlapping genes from the list of the top DEGs, but failed to reach threshold for significance, DEGs (*SEMA3*, replicated by Garbett et al. (2015a) and *IFIT3*, replicated after Segman et al. (2010)). Although we show that transcriptomes obtained from isolated cells provide results with higher replicability, increased risk of spontaneous modifications in gene expression during cell isolation process should be considered when interpreting the results.

Rather unexpectedly, the largest collection of replicated DEGs (9 out of 21) belongs to gene expression profiling study of dermal fibroblasts (Garbett, Vereczkei et al. 2015). Human dermal fibroblasts, due to their genetic and chemical stability during division in cell culture, are believed to be an experimental model for psychiatric research free of medication and lifestyle effects (Kalman et al., 2016) and have previously been used successfully for identifying cell cycle abnormalities in schizophrenia (Wang et al., 2010). This suggests that dermal fibroblasts may be fruitfully utilized as an accessible experimental model in depression gene expression research.

Comorbidities of depression

Depression is known for its high comorbidity with chronic somatic diseases, such as asthma, diabetes and cardiovascular disease. It may also occur in pregnancy and post pregnancy periods. Emerging research suggests that the comorbidity may arise due to overlapping mechanisms between different medical conditions. For this reason, we included and compared the overlap between the only two transcriptome studies investigating comorbid phenotypes with our list of replicated depression-related DEGs, (i) depressive asthma (Wang et al., 2015) and (ii) postpartum depression (Segman et al., 2010). Wang et al. (2015) hypothesized that asthma and depression were linked via overlapping molecules/pathways mediated by $CD4^+$ T-cells, and therefore isolated them and performed transcriptome profiling. Among 156 genes identified as differentially expressed between depressive asthma and healthy controls, we found that 6 genes (*ZNF333, PIK3R1, ZCCHC2, GNTAB, LMNA*, and *ABL1*) overlapped with the findings from other studies on pure MDD phenotype; among 73 DEGs between postpartum depression and healthy controls, 5 DEGs (*HIST1H4A, HIST1H4L, GSTT1, ASPM,* and *IFIT3*) overlapped with pure depression phenotype findings; moreover, one gene, *SERPING1*, showed overlap between depressive asthma and postpartum depression. These support the hypothesis of overlapping mechanisms involved in pathophysiology of depression and non-psychiatric diseases.

Stimulated blood cells challenge

Challenged blood cells potentially have a better ability to overcome the noise of variation in expression in basal blood, which is of critical importance in the case of such a heterogeneous phenotype as depression (Elowitz 2002). However, studies on stimulated cells, which are expected to show better signal-noise discrimination, showed the smallest number of replicated DEGs. Using *in vivo* dexamethasone stimulation of glucocorticoid

receptors (GR) in males aged 18-65 (18 MDD cases), Menke et al. (2012) were able to demonstrate the GR-mediated changes in gene expression between depressed patients and healthy controls. They suggested that altered GR functioning could potentially be used as a molecular biomarker of depression. However, only one gene (*SSH2*) out of 19 identified as a biomarker panel by Menke et al. (2012) was recently replicated in a larger cohort of depressed patients (Jansen et al., 2016).

Interestingly, one of the key GR binding proteins coded by the *STAT3* gene was significantly differentially expressed (upregulated) in both PBMCs (Belzeaux et al., 2012) and unstimulated whole blood (Jansen et al., 2016), but did not reach significance in the dexamethasone stimulated study.

The ex-vivo lipopolysaccharide (LPS) stimulation of whole blood suggested 12 top DEGs as a signature of depression (Spijker et al., 2010). However, only two genes, *LMNA* and *CAPRIN1,* were replicated by Wang et al. (2015) and Yi et al. (2012), respectively. Interestingly, *LMNA* was replicated in the depressive asthma phenotype study by Wang et al. (2015). Given that *LMNA* is associated with the most common type of cardiomyopathy, i.e. dilated cardiomyopathy (Lu et al., 2011), a disease of the heart muscle which primarily affects the left ventricle, a link between depression, asthma, and cardiovascular diseases is suggested.

Interplay between brain and periphery in depression

In an attempt to answer the question whether there is any interplay between the brain and periphery in depression at the replicated DEG level, we compared the lists of replicated DEGs for both brain and peripheral tissues. None of the replicated genes overlapped across both lists. This might suggest that there is no interaction between the CNS and periphery in depression at the replicated gene level. However, when we

compared the results from top disease and functional annotation, we observed that both lists were overrepresented with cardiovascular system abnormality terms. We further investigated this using Top Functions and Disease Network tool (IPA), and found that 25 out of 57 genes (43.9%) were replicated in the different brain areas in relation to depression (*GRIA1, GRIA2, GRIA4, GABRA1, GABRA5, GABRB1, GABRG2, GABRG1, GABRD, HTR2C, HTR2A, HTR1A, SLC1A2, SNAP25, SAT1, MAPT, JUN, NR3C1, CREB1, MOG, NTRK2, PTK2B, AQP4, ADRA2A, S100B*) also appeared to play a role in cardiovascular diseases, like coronary artery disease, peripheral vascular disease, hypertrophic response of cardiomyocytes, as well as permeability of vascular system and function of blood brain barrier (Figure 2.4). Furthermore, 6 out of 21 (28.6%) replicated depression-related DEGs in peripheral tissues (*STAT3, LMNA, MYH9, SERPING1, PIK3R1, ABL1*) were also involved in cardiovascular conditions, like dilated cardiomyopathy 1A, myopathy of heart, quality of cardiomyocytes, injury of heart, Slovenian type heart-hand syndrome, delay in initiation of leakage of capillary vessels, dilated cardiomyopathy with hypergonadotropic hypogonadism as well as permeability of blood vessels (Figure 2.5).

Figure 2.4. The 25 replicated genes identified as differentially expressed in depression at the genome-wide level in the brain and involved in the pathophisiology of cardiovascular diseases (Top Diseases and Functions Network analysis, IPA).

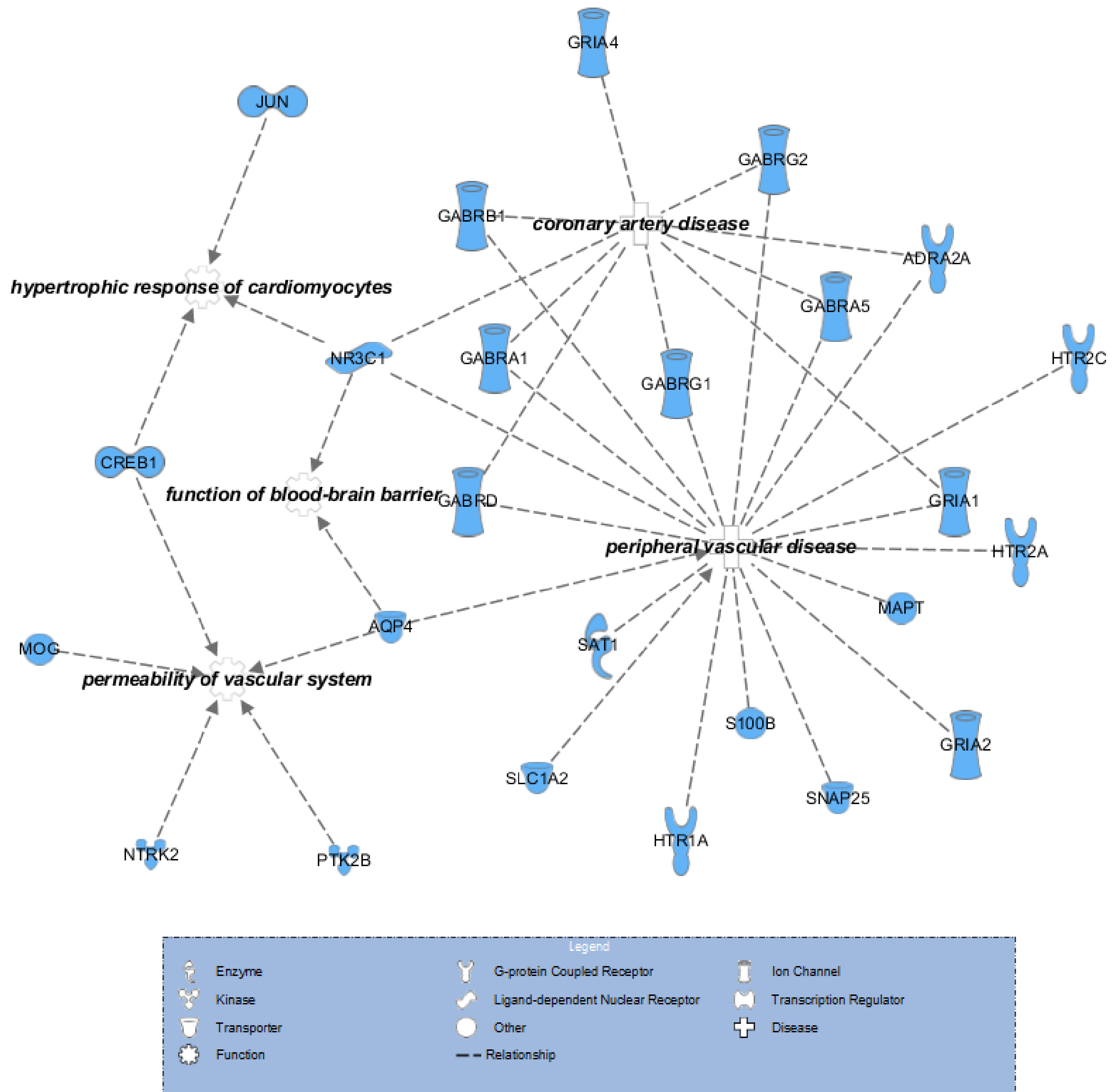
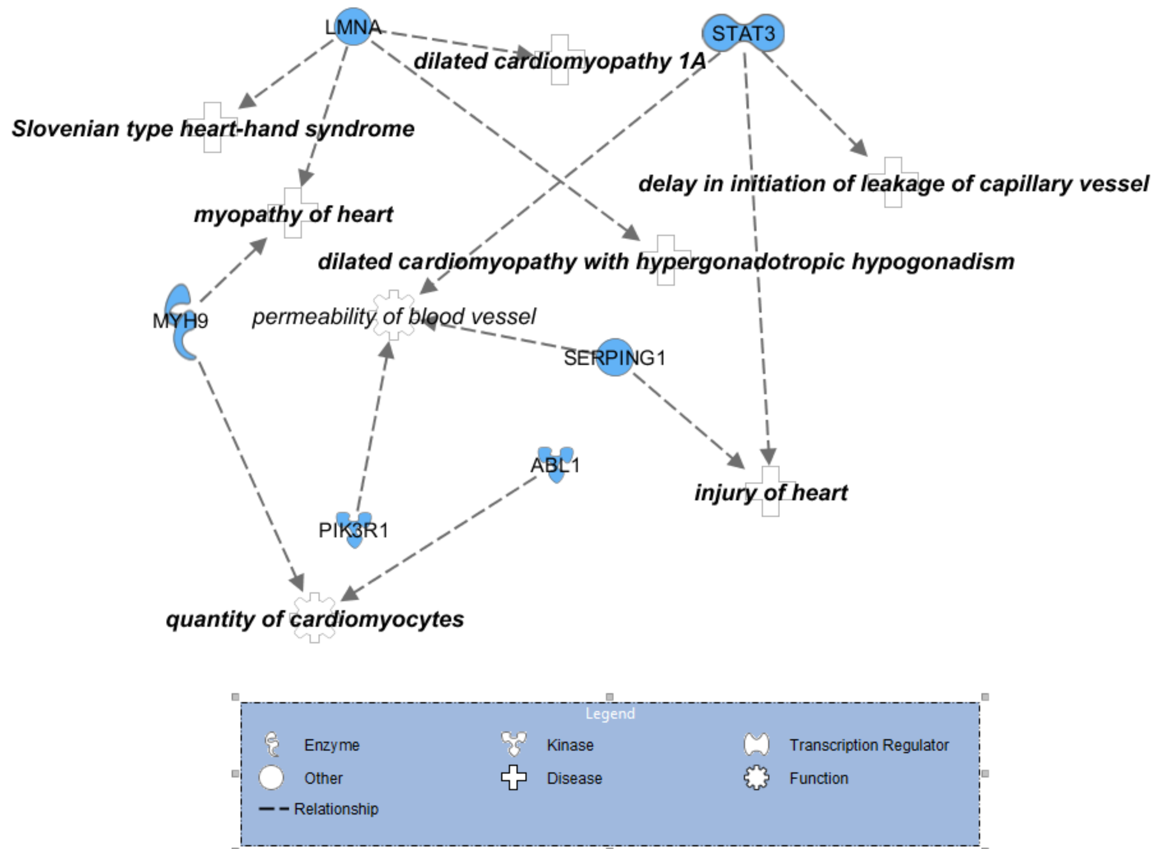


Figure 2.5. The 6 replicated genes identified as differentially expressed in depression at the genome-wide level in peripheral tissues and involved in the pathophisiology of cardiovascular diseases (Top Diseases and Functions Network analysis, IPA).



These findings support the hypothesis that depression and cardiovascular diseases may share molecular pathways (Bondy, 2007), and suggests that there are complex molecular interactions between brain cortical and subcortical areas and peripheral tissues in a depression/cardiovascular comorbid phenotype. These findings may have clinical value, as the link between the two conditions at the molecular level may potentially lead to the development of drugs that provide treatment for both cardiovascular conditions and depression.

Discussion

We reviewed the peripheral and brain genome-wide gene expression studies on depression. To the best of our knowledge this is the first attempt to systematically

catalogue all reported differentially expressed genes in depression. We observed 57 replicated differentially expressed genes (DEGs) across different brain regions (based on 15 studies) and 21 replicated DEGs in peripheral tissue (based on 10 studies). Twenty-eight of the brain DEGs were replicated in the same region. There was no overlap between the replicated genes in the brain and the periphery. Despite the replication, the directionality of dysregulation was highly inconsistent.

Our results suggest that brain transcriptomes are in better agreement across studies than transcriptomes derived from peripheral tissues. However, this might simply reflect the fact that brain tissues (both genome-wide and candidate genes) in relation to depression are better investigated than peripheral tissues.

The heterogeneity of the disease, dynamic nature of gene expression and the methodological differences between the studies all contribute to the observed low replication rate, which is the major obstacle for further progress in elucidating the mechanisms of depression. There are numerous technical factors that can potentially contribute to the observed inconsistencies across studies; however, this can be explained, at least partially, by biological factors. Given that mRNA abundance was measured in various cell types across multiple biological systems, the level and direction of gene expression may naturally differ by their functionality as different cell types perform different functions, and, therefore, the level of mRNAs can vary accordingly. Moreover, the state of disease changes over time, which is not necessarily reflected in a cross-sectional analysis at a single time point. A major question yet to be resolved is how a particular gene in a specific cell type contributes to the global picture of the disease and whether there is an interaction between the brain and periphery in depression.

There are several biological theories of the pathophysiology of depression, including altered hypothalamic–pituitary–adrenal axis **(**HPA) activity, monoamine deficiency,

neurotoxic and neurotrophic processes, reduced GABAergic activity, dysregulation of glutamate system, alteration of neuroimmune and cytokine activity, and impaired circadian rhythm (for review, see (Hasler, 2010)). Below we discuss how these findings relate to the major hypotheses of depression and suggest a new candidate to explain pathophysiological mechanisms of depression.

EVIDENCE SUPPORTING THE MAJOR HYPOTHESES OF DEPRESSION

Altered HPA axis activity

Glucocorticoids are involved in regulation of many organs and systems in the body, including various brain regions and molecular mechanisms, including monoaminergic neurotransmission and the immune and metabolic systems. Despite numerous findings suggesting that altered HPA activity plays a role in the pathogenesis of depression (Schatzberg et al., 2014, Vreeburg et al., 2009, Hardeveld et al., 2014), examination of the effect of pharmacological modulations of the neuroendocrine system as antidepressant therapy has been limited (Maric and Adzic, 2013). Our findings highlight the glucocorticoid receptor, *NR3C1* (see Table 2.1), which operates as a transcription factor in the regulation of gene expression (Lu and Cidlowski, 2005), as being dysregulated across the brain, i.e. downregulated in primary motor cortex (Sequeira et al., 2006) and upregulated in dorsolateral prefrontal cortex (BA8/9, BA46) (Sibille et al., 2004). However, inconsistency in directionally of dysregulation together with lack of brain region-specific replications suggests only partial support for altered HPA-axis activity and this hypothesis of depression.

Deficiency of monoamines

The monoamine-deficiency hypothesis states that the pathophysiological basis of

depression is a depletion of the neurotransmitters serotonin, norepinephrine or dopamine in the CNS. In this review, we show replicated evidence that dysregulation of serotonin receptors A1, 1B, 2A, 2C (*HTR1A, HTR1B, HTR2A, HTR2C)* predominantly in the dorsolateral prefrontal cortices (BA8/9, BA46) and hippocampus is involved in the pathophysiology of depression (Duric et al., 2013, Sibille et al., 2004). The *ADRA2A* gene, coding for alpha-2A adrenergic receptor, which localizes post-synaptically to noradrenergic terminals and promotes the function of norepinephrine, was consistently downregulated in the dorsolateral prefrontal cortex (BA8/9), pars orbitalis (BA47) (Sibille et al., 2004) and pars triangularis, part of the inferior frontal gyrus and part of Broca's area (BA45) (Klempan et al., 2009). While two studies were in agreement on directionality of dysregulation of *ADRA2A* across the brain, there is no brain area-specific replication so far. Although there are recent findings that a dopamine genetic risk score can predict depressive symptoms (Pearson-Fuhrhop et al., 2014), we cannot provide support that dopaminergic-related gene expression is dysregulated in depression. While almost all established antidepressant drugs target the monoamine system, pointing towards the monoamine-deficiency hypothesis as the most clinically relevant theory of depression, the resistance to these drugs, the delayed treatment effects and only partial support of this hypothesis at the level of replicated findings suggest that the dysregulation in monoamine system alone cannot explain the full spectrum of pathophysiological events in depression.

Altered neural plasticity and neurogenesis

There is consistent evidence that untreated depression leads to brain volume shrinkage in advanced age (Dotson et al., 2009) as well as in middle-aged depression patients (Grieve et al., 2013). Numerous findings have collectively proposed increased glucocorticoid and glutamatergic neurotoxicity and decreased neurotrophic factors and

neurogenesis as possible mechanisms explaining brain volume loss in depression. However, the understanding of molecular events leading to structural changes and neurodegeneration remains elusive. Brain-derived neurotrophic factor (*BDNF*) has attracted considerable interest as the major known regulator of synaptic plasticity in the brain. Although we did not find evidence for dysregulated transcription of *BDNF* mRNA in depressed patients, the *BDNF* receptor – the Tyrosine receptor kinase B protein coded by *NTRK2* – has been consistently found to be dysregulated across the cortex in 4 independent studies (see Table 1); upregulated in the anterior prefrontal cortex (BA10) (Malki et al., 2015) and hippocampus (Sequeira et al., 2007), and downregulated in the dorsolateral prefrontal cortex, pars orbitalis (BA47) (Sibille et al., 2004) and temporal cortex (BA21) (Aston et al., 2005). Another candidate for supporting the neurotrophic hypothesis of depression is fibroblast growth factor receptor 3, coded by *FGFR3*, which plays an important role in controlling cell growth and development. The *FGFR3* has been replicated in three studies, downregulated in the cortex (Kang et al., 2007, Evans et al., 2004) and the subcortical locus coeruleus (Bernard et al., 2011). Therefore, there is partial support for the neurodegenerative hypothesis of depression.

Reduced GABAergic activity

GABA is the major inhibitory neurotransmitter in the brain that modulates ongoing activity of neuronal networks. GABA re-uptake from the synaptic cleft is one the mechanisms to regulate GABA activity. Eight subunits for GABA receptors (*GABRA1, GABRA5, GARB1, GABRD, GABRG1, GABRG2, GABRR1*, and *GABBR2*) have been found to be consistently dysregulated, predominantly upregulated, across multiple brain areas including BA6, BA8/9, BA44, BA45, and BA46 (Table 1). *SLC6A1* - GABA transporter type 1, which functions to remove GABA from the synaptic cleft (Hirunsatit et al., 2009), has also been replicated as being downregulated in BA47 (Table 1). These findings mostly support the GABAergic hypothesis of depression. However, since the functionality

of GABA receptors is determined by specific subunit configurations (Anisman et al., 2012), the highly complex mosaic of possible interactions creates a myriad of possible effects on brain activity. What the relationships between GABA receptor subunits altered in depression are, and how the brain area-specific patterns of GABA receptor/transporter dysregulations are related to depression remain questions to be investigated in further research.

Dysregulation of glutamatergic system

Glutamate is the major excitatory neurotransmitter in the brain that regulates synaptic transmission and plasticity by activating ionotropic glutamate receptors, AMPA and NMDA. We found evidence that all four genes coding for AMPA receptor subunits (*GRIA1, GRIA2, GRIA3, GRIA4*) were replicably upregulated across multiple brain areas (see Table 1) with one exception – *GRIA3*, coding for AMPA receptor subunit 3, which was found to be upregulated in dorsolateral prefrontal cortex (BA46) by Choudary et al. (2005) and downregulated by Sequeira et al. (2009) and Klempan et al. (2009). The glutamate metabotropic 3 receptor (GRM3) has been replicated as being downregulated in temporal cortex (BA20) and BA47 in suicides with major depression (Sequeira et al., 2009, Klempan et al., 2009) The function of *SLC1A3* – glutamate aspartate transporter - is the termination of excitatory neurotransmission in CNS. The *SLC1A3* was downregulated in the locus coeruleus, known to be involved in the physiological response to stress (Bernard et al., 2011). Another glutamate transporter that clears the neurotransmitter from the extracellular space at synapses, which is necessary for proper synaptic activation and the prevention of neuronal damage from excessive activation of glutamate receptors - *SLC1A2*, has been replicated in the dorsolateral prefrontal cortices (BA8/9, BA47), anterior cingulate cortex (BA24) and locus coeruleus (Table 2.1). However, the directionality of *SLC1A3* dysregulation is inconsistent across different brain

regions as well as within the same brain area. Although these findings provide support for the glutamate system hypothesis of depression, the brain area-specific mechanisms of action and their effects on depressive behaviours require further investigation.

Disturbed neuroimmune and cytokine activity

There is growing evidence suggesting that cytokine-mediated neuroimmune disruptions contribute to the behavioural symptoms of depression (Loftis et al., 2010, McAfoose and Baune, 2009, Mills et al., 2013). Although we did not find support for the most promising candidates of the cytokine hypothesis of depression, like *IL1B, IL6*, and *TNFα* at the level of replicated genome-wide gene expression findings, it is worth noting that many genes from our replicated genes list are known to be involved in various aspects of immune responses. Using GO search (ImmPort database) we identified that (1) four replicated genes - *FGFR3, ENPP2, PTP4A2,* and *CREB1* – are involved in the innate immune response; (2) three genes – *CREB1, MOG,* and *JUN* are heavily involved in various toll-like receptor signalling pathways; (3) *LEPR,* which belongs to the gp130 family of cytokine receptors, is involved in cytokine-mediated signalling. These collectively suggest that altered immunological activation may play a role in depression via various pathways, possibly interacting with other pathological mechanisms of depression.

Dysregulation of melatonergic system

The melatonergic hypothesis of depression states that dysregulation of melatonin plays a principal role in sleep disturbances in depressed patients. Since serotonin is the precursor of melatonin, two systems – melatonergic and monoaminergic - become tightly tied together, providing new avenues for future research. Although this review cannot directly support the melatonergic hypothesis of depression, as we did not observe differential expression of membrane MT1/MT2 melatonin receptors, through which

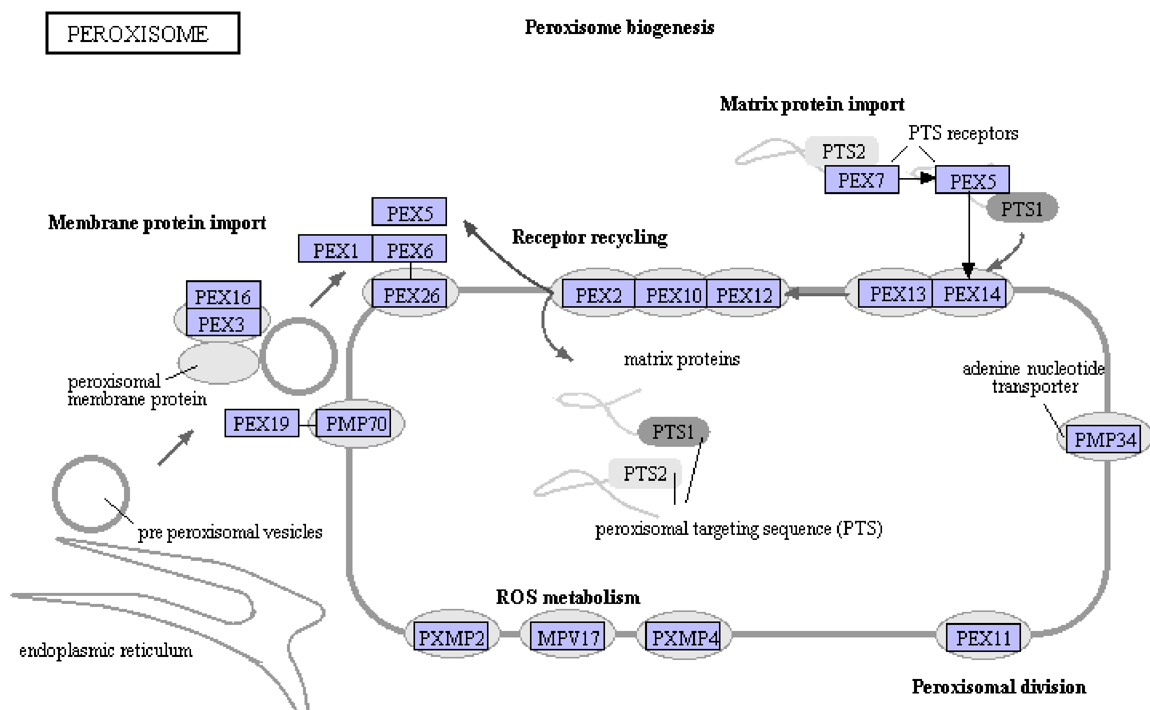
melatonin exerts its action, nor for specific clock genes *Per1* and *Per2*, we can provide indirect evidence for this hypothesis. Transcription factor *CREB1* - CAMP responsive element binding protein 1, which mediates norepinephrine-related mechanisms of melatonin secretion (Maronde et al., 2011), was dysregulated in depression compared to controls across the brain, being downregulated in the dorsolateral prefrontal cortex (BA8/9, BA47) (Sibille et al., 2004) and upregulated in the anterior prefrontal cortex (BA10) (Tochigi et al., 2008). However, it is not known why *CREB1* is dysregulated in opposite directions in different brain areas and how it is related to depression.

Novel candidate of pathophysiology of depression

Replicated findings at the genome-wide gene expression level provide partial support for all major theories of depression. This clearly argues against a single theory and suggests that depression is etiologically a highly heterogeneous disorder. Moreover, replication of the genes that are not yet known to have a direct relationship with any of the existing theories of depression may navigate towards new molecular players in understanding the biology of depression. One of the most exciting findings is dysregulation of *PXMP2* in dorsolateral prefrontal, anterior cingulate and temporal cortices. Peroxisomal membrane protein 2, coded by *PXMP2* gene, is a channel-forming protein in a mammalian peroxisomal membrane (Figure 2.6). Peroxisomes are cell organelles primarily involved in lipid metabolism and reduction of reactive oxygen species (ROS), specifically hydrogen peroxide, which are critical for the normal functioning of the brain (Antonenkov and Hiltunen, 2012, Wanders and Waterham, 2006). Although oxidative stress been previously found to be implicated in depression, which is partially reflected by our finding highlighting replicated downregulation of *LEPR* and *S100B* in the brain and periphery, to the best of our knowledge, neither peroxisome functioning in general nor dysregulation of *PXMP2* in particular, have ever been investigated in relation to mood disorders, including

depression. Taking into account findings pointing towards shared pathways between metabolic disorders and depression (Foley et al., 2010) and robust replication of this gene at the genome-wide gene expression level in depression, we suggest that dysregulation of *PXMP2* may play a role in depressive disorder via the peroxisome lipid and ROS metabolism pathway.

Figure 2.6. Peroxisome and the role of *PXMP2* in lipid and reactive oxygen species (ROS) metabolism (KEGG pathway).



Limitations

Although replication of findings is a way to systematise the most robust findings at the single gene level, the approach itself can be viewed as a limiting factor for the findings that have not yet been replicated. Non-replicated results, especially those of high significance, are likely to have a biological value but cannot be included in this review of replicated findings. To overcome this limitation, more studies are needed.

In the current review, we examined replication of gene expression results in the periphery

and the brain. This approach has an obvious limitation as there are more parameters to consider, like phenotype definition (e.g. age of onset, diagnosis, method of recruitment, use of medication), sample characteristics including race, age and sex, microarray platform and statistical algorithms used. To include all those variables, a more hierarchical approach is needed.

Another important limitation of this study is missing or incomplete information reported in the article. Given that we examined the genes derived from differential expression analysis only, we were expecting to find the full lists of DEGs either in the main body text or in the supplement materials. However, reporting only the top DEGs (with the lowest p-value) is a common practice for transcriptome studies. A related limitation is that the choice DEGs was based on an arbitrary p-value threshold selected by the authors, which may differ between studies. To overcome these limitations authors are encouraged to provide the complete list of differentially expressed genes along with corresponding fold-change and p-values.

The small sample sizes are the common limiting issue for the microarray and RNA-seq studies due to the cost and difficulties in obtaining samples, particularly for brain studies. Global collaborative efforts at the consortia level in data collection, advancements in data analysis and good communication practices may help us to better understand the pathophysiological mechanisms of depression.

Future directions

Given that depression is a complex disease with overlapping activity across many systems, the key pathophysiological mechanisms can be obscured when investigated by measuring the relative difference in mRNA abundance between cases and controls. Growing interest in investigation using coexpression (Zhang and Horvath, 2005) and

differential coexpression (Zheng et al., 2014) network analyses, which measure the coordinated coexpression patterns between the genes, can provide a far more detailed and complete picture of the pathophysiology of depression. The observation that differentially expressed genes reside at the periphery of the co-expression networks in depression (Gaiteri et al., 2010a, Gaiteri et al., 2014) suggests that differential expression technique is not an optimal method to detect the key regulatory genes. This observation is consistent with the heterogeneous nature of disease and may partially explain the limited therapeutic effect of currently available antidepressant drugs, as pharmacological modulation of one or even a set of genes with low connectivity are likely to have minimal effect on key regulators. Network-based methodologies as part of systems biology approach are promising innovations that can provide new insights into molecular processes underlying the pathophysiology of depression, which potentially can lead to development of clinical biomarkers required for diagnosis, drug development and improved treatment of depression.

Supplementary data associated with this article, can be found in the online version, at http://dx.doi.org/10.1016/j.neubiorev.2016.08.018

FROM A SINGLE GENE TO A NETWORK OF COMPLEX INTERACTIONS

In the first chapter, the gene expression signature of depression based on a large body of previously obtained findings in both brain and peripheral tissues covering 16 brain regions and five cell types from the peripheral nervous system was examined. We found, that out of 1,334 genes reported as differentially expressed across 25 transcriptome studies of depression, only 57 genes in the brain and 21 in the peripheral tissues were replicated, highlighting large inconsistencies across the studies. An overlap in genetic expression between the brain and peripheral tissues was also observed which strongly implicated a link between depression and cardiovascular disease.

This study suggests that depression is associated with dysregulated expression of multiple genes (at least several hundreds), supporting multiple theories of the underlying biological mechanisms of depression. In-silico functional characterisation of 'replicated' genes using Ingenuity Pathway Analysis (IPA) indicated that these genes are involved in many biological functions, such as cell cycle, cellular function and maintenance, cellular death and survival. Given the nature of biological systems, these individual genes are unlikely to act in isolation: instead, they interact with each other forming complex molecular networks. These interacting genes can be involved in various biological processes, therefore affecting multiple molecular pathways, the dysregulation of which may lead to depression. Examining the global network of co-expressed genes at the genome-wide level will help in understanding which known molecular pathways are dysregulated within depression therefore providing the insights into the molecular underpinnings of the disorder.

In the next chapter, we explore which patterns of co-expressed genes are associated with depression and whether these patterns are functionally meaningful. This work consists of four major steps: (1) investigation into the network structure of peripheral blood

transcriptome using an unsupervised data mining technique, Weighed Gene Co-expression Network Analysis (WGCNA), (2) identification of which clusters/modules of tightly co-regulated genes are associated with MDD disease status, (3) examination of which molecular pathways appear to be altered in depression using in-silico functional characterisation of MDD-related clusters, and (4) replication of the findings on an independent cohort.

This study has the potential to improve our understanding of biological underpinnings of depression, pinpoint specific biological markers and suggest which molecular pathways are dysfunctional in depression.

Statement of Authorship

Title of Paper	Weighed gene co-expression network analysis (WGCNA) of peripheral blood transcriptome in recurrent MDD in older adults	
Publication Status	☑ Published ☐ Submitted for Publication	☐ Accepted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	CIOBANU, L. G., SACHDEV, P. S., TROLLOR, J. N., REPPERMUND, S., THALAMUTHU, A., MATHER, K. A., COHEN-WOODS, S., STACEY, D., TOBEN, C., SCHUBERT, K. O. & BAUNE, B. T. 2018. Co-expression network analysis of peripheral blood transcriptome identifies dysregulated protein processing in endoplasmic reticulum and immune response in recurrent MDD in older adults. *Journal of Psychiatric Research.*	

Principal Author

Name of Principal Author (Candidate)	CIOBANU, L. G.		
Contribution to the Paper	Conceived of the presented research concept and design, analysed clinical data to derive recurrent depression subtype, generated and pre-processed gene expression data, performed statistical analysis, interpreted the results, wrote the manuscript and submitted to the journal, finalised manuscript based on reviewers feedback.		
Overall percentage (%)	85%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	20.05.2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	SACHDEV, P. S.		
Contribution to the Paper	• sMAS and OATS cohort study design and data collection • Critical revision of the manuscript • Discussed the results and contributed to the final manuscript		
Signature		Date	11 June 2019

Name of Co-Author	TROLLOR, J. N.		
Contribution to the Paper	• sMAS and OATS cohort study design and data collection • Critical revision of the manuscript • Discussed the results and contributed to the final manuscript		
Signature		Date	11/06/2019

Name of Co-Author	REPPERMUND, S.		
Contribution to the Paper	• sMAS and OATS cohort study design and data collection • Critical revision of the manuscript		
Signature		Date	11/06/2019

Name of Co-Author	THALAMUTHU, A.		
Contribution to the Paper	• sMAS cohort study design and data collection • Critical revision of the manuscript		
Signature		Date	12/06/2019
Name of Co-Author	MATHER, K. A.		
Contribution to the Paper	• sMAS and OATS cohort study design and data collection • Critical revision of the manuscript		
Signature		Date	6/6/19
Name of Co-Author	COHEN-WOODS, S.		
Contribution to the Paper	• Critical revision of the manuscript • Contributed to the final manuscript		
Signature		Date	29/05/2019
Name of Co-Author	STACEY, D.		
Contribution to the Paper	• Inspired on analytical approach and verified the analytical methods • Critical revision of the manuscript		
Signature		Date	31/05/2019
Name of Co-Author	TOBEN, C.		
Contribution to the Paper	• Critical revision of the manuscript • Helped shape the final manuscript		
Signature		Date	10.05.2019
Name of Co-Author	SCHUBERT, K. O.		
Contribution to the Paper	• Critical revision of the manuscript • Helped define the writing style		
Signature		Date	21.5.19
Name of Co-Author	BAUNE, B. T.		
Contribution to the Paper	• Supervised the project • Provided critical feedback to study design and methodology • Provided critical feedback to the first draft of the manuscript • Discussed the results and contributed to the final manuscript • Final decision on manuscript submission		
Signature		Date	16.06. 2019

CHAPTER 3

Co-expression network analysis of peripheral blood transcriptome identifies dysregulated protein processing in endoplasmic reticulum and immune response in recurrent MDD in older adults

Journal of Psychiatric Research 107 (2018) 19–27

Contents lists available at ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/jpsychires

Co-expression network analysis of peripheral blood transcriptome identifies dysregulated protein processing in endoplasmic reticulum and immune response in recurrent MDD in older adults

Liliana G. Ciobanu[a], Perminder S. Sachdev[b], Julian N. Trollor[b,c], Simone Reppermund[b,c], Anbupalam Thalamuthu[b], Karen A. Mather[b,f], Sarah Cohen-Woods[d], David Stacey[e], Catherine Toben[a], K. Oliver Schubert[a,g], Bernhard T. Baune[a,h,*,1]

[a] Discipline of Psychiatry, Adelaide Medical School, University of Adelaide, South Australia, Australia
[b] Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, UNSW Sydney, New South Wales, Australia
[c] Department of Developmental Disability Neuropsychiatry, School of Psychiatry, UNSW Sydney, New South Wales, Australia
[d] School of Psychology, Flinders University, South Australia, Australia
[e] MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK
[f] Neuroscience Research Australia, Randwick, Australia
[g] Northern Adelaide Local Health Network, Mental Health Services, Lyell McEwin Hospital, Elizabeth Vale, South Australia, Australia
[h] Department of Psychiatry, Melbourne Medical School, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Parkville VIC, 3010, Australia

Abstract

The molecular factors involved in the pathophysiology of major depressive disorder (MDD) remain poorly understood. One approach to examine the molecular basis of MDD is co-expression network analysis, which facilitates the examination of complex interactions between expression levels of individual genes and how they influence biological pathways affected in MDD. Here, we applied an unsupervised gene-network based approach to a prospective experimental design using microarray genome-wide

gene expression data derived from the peripheral whole blood of older adults. We utilised the Sydney Memory and Ageing Study (sMAS, N=521) and the Older Australian Twins Study (OATS, N=186) as discovery and replication cohorts, respectively. We constructed networks using Weighted Gene Co-expression Network Analysis (WGCNA), and correlated identified modules with four subtypes of depression: single episode, current, recurrent, and lifetime MDD. Four modules of highly co-expressed genes were associated with recurrent MDD (N=27) in our discovery cohort, with no significant findings for a single episode, current or lifetime MDD. Functional characterisation of these modules revealed a complex interplay between dysregulated protein processing in the endoplasmic reticulum (ER), and innate and adaptive immune response signalling, with possible involvement of pathogen-related pathways. We were underpowered to replicate findings at the network level in an independent cohort (OATS), however; we found a significant overlap for 9 individual genes with similar co-expression and dysregulation patterns associated with recurrent MDD in both cohorts. Overall, our findings support other reports on dysregulated immune response and protein processing in the ER in MDD and provide novel insights into the pathophysiology of depression.

INTRODUCTION

Major depressive disorder (MDD) is a leading cause of disability worldwide and is one of the major contributors to the overall global burden of disease spanning all age groups (Ferrari et al., 2013, Whiteford et al., 2015, Vos et al., 2016). While MDD is relatively uncommon among older adults, epidemiological studies suggest that clinically significant depressive symptoms affect between 7% and 49% of community-dwelling older adults, substantially affecting the quality of life in later years (Riedel-Heller et al., 2006). MDD is highly heterogeneous with regards to its clinical phenotypes (Fried, 2017), course of symptoms (Posternak et al., 2006), responses to treatment (Uher and Pavlova, 2016),

and longer-term functional outcomes (Prisciandaro and Roberts, 2008). Major biological hypotheses of MDD have included altered hypothalamic–pituitary–adrenal (HPA) axis activity (Pariante and Lightman, 2008), deficiency of monoamines (Hirschfeld, 2000), altered neural plasticity and neurogenesis (Jacobs et al., 2000), oxidative stress (Maes et al., 2011), dysregulation of GABAergic, glutamatergic (Luscher et al., 2010, Pan et al., 2018) and melatonergic (Srinivasan et al., 2009) systems, and dysregulated neuroimmune pathways (Wohleb et al., 2016, Leday et al., 2018).

High-throughput methodologies which allow screening for a large number of biological substrates for molecular differences between MDD patients and healthy controls in hypotheses-free paradigms are a promising avenue towards improving our understanding of the disorder. Several genome-wide association studies (GWAS) in MDD have indicated that the genetic architecture of MDD is complex, with many polymorphisms of small effect contributing to the clinical phenotype in middle-aged adults (Cai et al., 2015, Okbay et al., 2016, Hek et al., 2013, Wray et al., 2017, Hyde et al., 2016) as well as in late-life depression (Tsang et al., 2017). Studying global gene expression is an emerging and continually growing field that can shed light on the molecular underpinnings of depression that can lead to the development of biology-based diagnostic tools and novel pharmacological therapies. A recent systematic review of 25 transcriptome studies found that over 1200 genes have been reported as differentially expressed in MDD compared to controls in the brain, and peripheral, tissues (Ciobanu et al., 2016). Replicability of these findings, however, has been minimal, which may be attributed to differences in methodological and statistical applications, small sample sizes, false positives, and the inherent dynamic and cell-specific nature of gene expression. In addition, clinical heterogeneity has been recognized as a major limiting factor for robust characterisation of gene expression alterations in MDD. For example, the first RNA sequencing study of 463 lifetime MDD cases, consisting of a mixture of individuals with current and remitted

MDD, found no differentially expressed genes between cases and controls (Mostafavi et al., 2014). In contrast, stratification of the phenotype into “current” and “remitted” MDD yielded 18 genes differentially expressed between control and current MDD groups (FDR<0.05) (Jansen et al., 2016). These findings have yet to be replicated. Thus, studying transcriptomic patterns among more homogeneous subgroups of MDD patients has the potential to improve the identification of a biomolecular signature of depression.

A challenge in interpreting gene expression data is that ‘candidate’ genes do not function in isolation but rather interact in complex networks which can affect clinical phenotypes. Weighted gene co-expression network analysis (WGCNA) is a hypothesis-free systems biology approach that identifies ‘modules’ of co-regulated, and therefore functionally related, genes in a given phenotype (Langfelder and Horvath, 2008), extending classic bivariate approaches.

In this study, we aimed to investigate the relationship between global gene co-expression profiles and MDD subgroups. Utilising a 6-year prospective community-dwelling sample of relatively healthy elderly people from the sMAS, we applied WGCNA, and explored the correlation of co-expressed modules with four phenotypes: (a) a lifetime diagnosis of MDD, (b) a single episode of MDD diagnosed during the study, (c) current episode of MDD, and (d) recurrent MDD diagnosed during the study. We then sought to replicate our findings in a second, independent cohort drawn from the Older Australian Twins Study (OATS), Australia.

METHODS

Discovery cohort

The Sydney Memory and Aging Study (sMAS) was initiated in 2005 to examine the clinical

characteristics and prevalence of mild cognitive impairment and related syndromes, including depression, in a relatively healthy aged population aged 70-90 years at recruitment (N=1,037) (Sachdev et al., 2010). The phenotype data were collected at four time points with 2-year intervals between assessments. Blood samples for gene expression analyses were collected at Wave 4 (N=521), six years after baseline data collection.

Replication cohort

To conduct the replication analyses we utilised the Older Australian Twins Study (OATS) (N=623).The primary aim of OATS is to investigate healthy brain ageing in older twins (65+ years) (Sachdev et al., 2009). Depression data were collected at three time points with 2-year intervals between assessments. Blood samples for gene expression analyses were collected at Wave 3, four years after baseline (N=186).

Informed consent was obtained for all participants and study procedures were explained prior to study commencement.

MDD definition

MDD was assessed by two well-validated self-report, and two clinical interview-based, measures of depression including the Geriatric Depression Scale (GDS-15) (cut-off ≥ 6) (Yesavage et al., 1982), the Patient Health Questionnaire (PHQ-9) (cut-off ≥ 10) (Kroenke K and R., 2002), the Neuropsychiatric Inventory (NPI) (depression sub-scale) (Cummings et al., 1994), and the Mini International Neuropsychiatric Interview (M.I.N.I.) (Sheehan et al., 1998). To make use of all measures available, a “minimum by two” approach for defining MDD was employed. Accordingly, MDD cases were defined as participants who were identified as clinically depressed by at least two of the above depression measures at any one study Wave, according to DSM-IV criteria. Using this approach, four subgroups

of depression were formed: a) lifetime diagnosis of MDD, b) a single episode of MDD, c) current diagnosis of MDD and d) recurrent MDD (details in Appendix A).

Gene expression data acquisition and pre-processing

Total RNA from whole blood collected in PAXgene tubes after overnight fasting was extracted using the PAXgene Blood RNA System (PreAnalytiX, QIAGEN). The RNA samples with RIN ≥ 6 were used in subsequent analyses (Gallego Romero et al., 2014). Purification, amplification, labelling, and microarray hybridization were carried out using the Illumina Whole-Genome Gene Expression Direct Hybridization Assay System HumanHT-12 v4 (Illumina Inc., San Diego, CA, USA) according to standard manufacturer protocols. Quality control (QC) and pre-processing of raw gene expression intensity values extracted from GenomeStudio (Illumina) were performed within the R statistical environment. The pre-processing steps included: (1) background correction by Maximum Likelihood Estimation (MLE) algorithm using Model-based Background Correction for Beadarray (MBCB) R package, (2) the Variance-Stabilising Transformation (VST), (3) quantile normalisation, and (4) two-stage filtering by detection p-value ($p<0.01$ in ≥50% of samples) and by coefficient of variation under threshold of 0.01. Data were adjusted for batch and RINs effects using Empirical Bayes-moderated linear regression implemented in *empiricalBayesLM* function (WGCNA package); control for non-biological latent noise was performed using *sva* function (Leek et al., 2012) (details in Appendix A).

Co-expression network analysis

WGCNA was performed with the WGCNA R package, as previously described (Langfelder and Horvath, 2008). In short, this method selects the threshold for constructing the network based on the scale-free topology of gene co-expression networks. Using biweight mid-correlation, which is a robust alternative of Pearson's

correlation (Langfelder and Horvath, 2012), we computed the networks for several thresholds and selected the threshold β=6, which led to a network with scale-free topology. The network consisted of multiple modules of functionally related genes on the basis of their expression patterns. Module eigengenes (MEs), i.e., the first principal components of the standardized gene expression profile of a given module, were tested for correlation with binary measures of depression subtypes versus the rest of the cohort. The MEs were used to define measures of module membership (MM) by correlating the gene expression profile of each gene in a module with the ME of a given module. The gene with the highest MM was defined here as the top hub gene of a given module. Gene significance (GS) was computed as hybrid robust-Pearson's correlation (Langfelder and Horvath, 2012) between gene expression and MDD status. Module Membership (MM) vs. Gene Significance (GS) measure was used to assess how representative each gene is of the module and its importance for MDD.

Cell-type specific gene expression

We examined whether known cell type-specific markers of expression (Aran et al., 2017) were driving depression-related gene expression signatures by calculating relative proportions of cell type-specific markers in defined MDD-related modules (details in Appendix A).

Functional analyses

To functionally characterise modules of interest we used two strategies: (1) examined the significance of pathway enrichment in a gene list of each module with a modified Fisher's exact test (FDR < 0.5) using the Database for Annotation, Visualisation and Integrated Discovery (DAVID) (Huang da et al., 2009b, Huang da et al., 2009a) with the genes which were included in the WCGNA as background, and (2) performed Signalling Pathway Impact Analysis (SPIA) (Tarca et al., 2009) implemented in the SPIA R package. SPIA uses the information from differentially expressed genes and their fold

changes (derived using *limma* R package (Ritchie et al., 2015) for each module separately), as well as pathways topology in order to assess the significance of the pathways. We used the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) database to run the SPIA algorithm. These functional analyses were performed for the modules correlated with recurrent MDD phenotype only (details in Appendix A)

Replication study

Data from 186 unrelated OATS participants (one twin from each pair, randomly selected) were used for replication analyses. To assess comparability between sMAS and OATS gene expression datasets, we computed Pearson's correlations between ranked expressions and ranked connectivity for probes that were expressed in both data sets (Miller et al., 2010). To conduct replication analyses we (1) extracted the OATS data for the transcripts that represented the sMAS depression-related modules, (2) computed MEs for these pre-constructed modules in the replication cohort, (3) correlated these MEs with a binary phenotype of recurrent MDD, and (4) tested for statistical significance of the overlap between two groups of recurrent MDD-correlated genes in two datasets (details in Appendix A).

RESULTS

Cohort characteristics

The sMAS sample consisted of 521 individuals aged over 76 (age range 76.4-95.9, mean age 83.7±4.5), 255 males and 266 females. Utilising the prospective nature of this study, four subgroups of MDD were defined. Statistical tests of independence showed that there were no differences between cases and controls for age, sex or BMI in neither of MDD patient subgroups (Table 3.1).

Table 3.1. Summary of MDD patient subgroups and statistical test of independence for Age, Sex and BMI between cases and controls for MDD patient subgroups in sMAS cohort.

MDD patient subgroups	Definition	Number of MDD cases (%)	Number of controls (%)	Total sample size	Age (t-test)	Sex (X^2)	BMI (t-test)
Lifetime diagnosis of MDD	Diagnosis of MDD in at least one Wave	83 (15.9)	438 (84.1)	521	t(519)=1.02, p=0.3	X^2(3, N=521)=2.21, p=0.13	t(519)=0.9, p=0.3
Single episode of MDD	Diagnosis of MDD in single Wave only	56 (10.8)	465 (89.2)	521	t(519)=0.9, p=0.3	X^2(3, N=521)=3.29, p=0.07	t(519)=0.8, p=0.4
Current MDD	Diagnosis of MDD at Wave 4	49 (9.4)	472 (90.6)	521	t(519)=1.7, p=0.07	X^2(3, N=521)=2.23, p=0.13	t(519)=1.4, p=0.2
Recurrent MDD	Diagnosis of MDD in two or more Waves	27 (5.2)	494 (94.8)	521	t(519)=0.1, p=0.9	X^2(3, N=521)=0.09, p=0.8	t(519)=0.7, p=0.5

Weighted gene co-expression network analysis

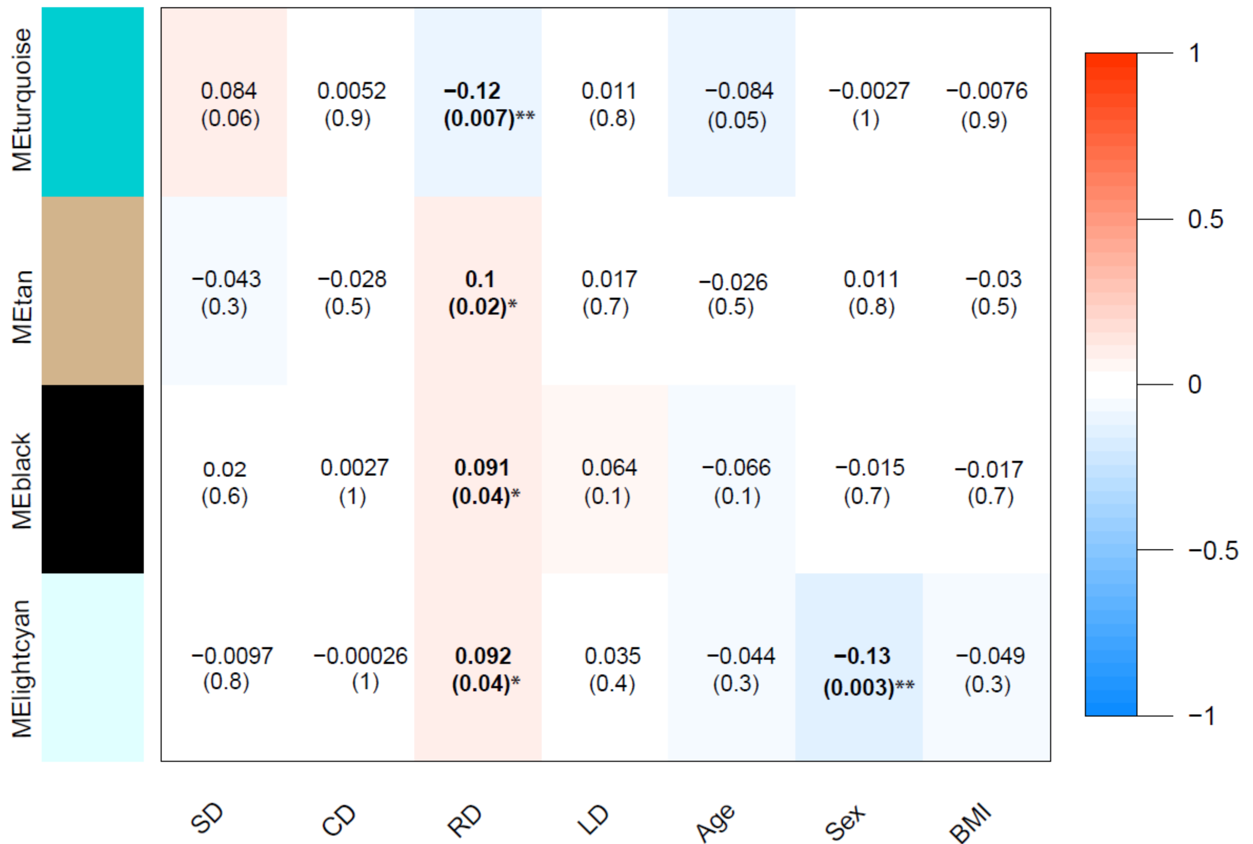
After pre-processing, 11 018 probes corresponding to 9 041 genes were used for downstream analyses. Using WGCNA we constructed a co-expression network for 521 samples and identified 24 distinct modules of various sizes (details in Appendix A).

Four gene co-expression modules are associated with recurrent MDD, but no module association was found for lifetime, single episode or current MDD

No association between any of the identified module eigengenes and lifetime, single episode or current MDD was observed. However, the eigengenes of four modules (denoted by colour) were significantly associated with recurrent MDD: 1) turquoise (r = -0.12, p = 0.007), 2) tan (r = 0.1, p = 0.02), 3) black (r = 0.09, p = 0.04), and 4) lightcyan (r = 0.09, p=0.04) (Figure 3.1). However, only turquoise module remained significant after

using Benjamini-Hochberg procedure at 0.2 FDR for 24 tests. The lightcyan module was also associated with sex ($r = -0.13$, $p = 0.03$). The MEs of turquoise, tan, and black modules were not associated with age, sex, or BMI (Figure 1).

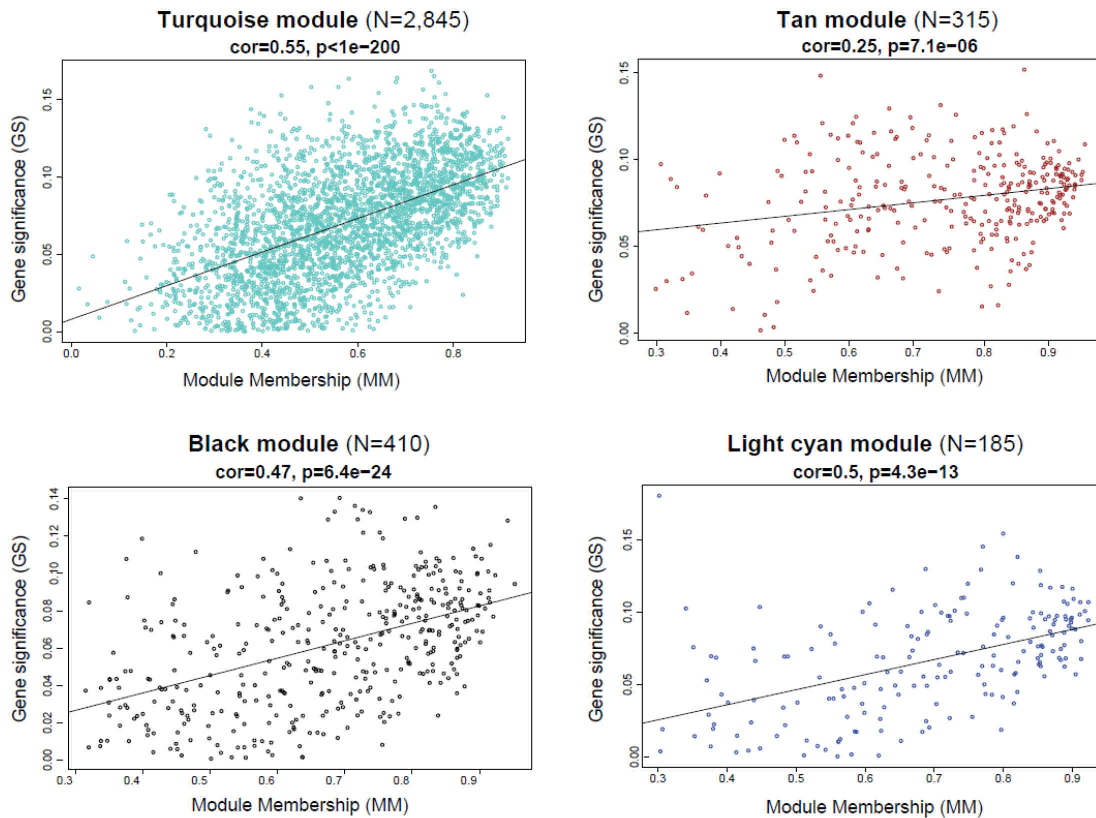
Figure 3.1. Heatmap plot of correlations between the module eigengenes (ME) of Turquoise, Tan, Black, Lightcyan modules and single episode of depression (SD), current (CD), recurrent (RD), and lifetime depression (LD), age, sex and BMI. Associated p-values are indicated in parenthesis with significant values bolded.



Relationship between MM and GS for recurrent MDD

We found a highly significant positive correlation between MM and GS measures for all four modules, indicating that those individual genes that were statistically representative of their modules were also important for recurrent MDD (Figure 3.2).

Figure 3.2. Scatterplots depicting correlations between Module Membership (MM) and Gene Significance (GS) for Turquoise, Tan, Black, and Lightcyan modules, eigengenes of which were associated with recurrent MDD in older adults.



In total, 1 241 probes were nominally associated with recurrent MDD across four modules (GS, $p < 0.05$); 761 probes remained significant after transcriptome-wide correction for multiple testing (FDR at 0.2) (Online Supplementary Tables S1, S2, S3, S4).

Immune cells-specific markers of expression for recurrent MDD-related modules

We defined gene expression signatures for six immune cell types – B cells, CD4+ T cells, CD8+ T calls, monocytes, neutrophils and NK cells – using in-silico estimation method (Aran et al., 2017). We found that turquoise module is likely to be influenced by these immune cell types; only negligible number of cell-specific markers was found in tan, black

and lightcyan modules.

The effect of smoking on the findings

To examine the potential effect of smoking on our results, we checked whether our modules of interest contain any genes the expression of which can be effected by smoking (Vink et al., 2017, Huan et al., 2016). We found that the Turquoise module contained several smoking-related genes (*CLDND1*, MM=0.78; *MTSS1*, MM=0.5; *PASK*, MM=0.3). However, we observed that these genes were not associated with recurrent MDD status and had moderate to low MM value; none of the hub genes (with MM>0.9) checked across four modules have been previously found associated with smoking, suggest that smoking is unlikely to have a large effect of these findings.

Enrichment and signalling impact pathway analyses

Enrichment analyses highlight downregulation of genes involved in protein processing in Endoplasmic Reticulum (ER) in recurrent MDD

Using DAVID we identified the biological pathways that were significantly over-represented in each of the four modules. The two most relevant pathways associated with turquoise module were: *protein processing in ER* (FDR = 2.8e-06) and *COPII (Coat Protein 2) Mediated Vesicle Transport* (FDR = 2.7e-06). We also found significant enrichment for *mRNA Splicing, antigen processing: ubiquitination & proteasome degradation, ubiquitin mediated proteolysis*, and *macroautophagy* pathways in turquoise module (Table 3.2). Given that over 79% of transcripts in turquoise module were downregulated (2250 downregulated, 595 upregulated), and the correlation between ME and recurrent MDD was negative (r = -0.12, p = 0.007) we concluded that downregulation of turquoise module may play an important biological role in recurrent MDD. The black module was enriched with the *hemoglobin's chaperone* pathway (FDR = 0.008). There

was no pathway enrichment for tan and lightcyan modules at the FDR ≤ 5% level.

Table 3.2. Pathway enrichment analysis on the modules associated with recurrent MDD in older adults

Database: Pathway ID	Pathway name	NP	NM	FE	PValue	FDR
Turquoise module (N=2,845)						
KEGG: hsa04141	*Protein processing in endoplasmic reticulum*	189	51	2.40	2.1E-09	2.8E-06
REACTOME: R-HSA-204005	*COPII (Coat Protein 2) Mediated Vesicle Transport*	68	30	3.34	1.7E-09	2.7E-06
REACTOME: R-HSA-72163	*mRNA Splicing - Major Pathway*	180	48	2.08	8.5E-07	0.001
REACTOME: R-HSA-983168	*Antigen processing: Ubiquitination & Proteasome degradatio*	309	71	1.75	1.9E-06	0.003
KEGG: hsa04120	*Ubiquitin mediated proteolysis*	155	38	2.18	4.3E-06	0.006
REACTOME: R-HSA-1632852	*Macroautophagy*	68	23	2.64	2.0E-05	0.03
Tan module (N=315)						
REACTOME: R-HSA-3928665	*EPH-ephrin mediated repulsion of cells*	49	5	7.60	0.004	5.15
Black module (N=410)						
BIOCARTA: h_ahspPathway	*Hemoglobin's Chaperone*	15	6	18.58	7.2E-06	0.008
Lightcyan module (N=185)						
REACTOME: R-HSA-139915	*Activation of PUMA and translocation to mitochondria*	9	2	36.43	0.05	4.49

Pathway enrichment analyses were performed using DAVID. As a background gene set, we utilised 9 041 genes used for network construction.
NP – the number of genes in the pathway, NM – the number of module genes featured in the pathway, FE – fold enrichment

SPIA identifies dysregulation of biological pathways involved in innate and adaptive immune response, protein processing in ER and host-defence response to infectious pathogens in recurrent MDD

To further detail how dysregulation of individual genes within each module affects specific pathways we selected module probes with GS p<0.05 (963 in turquoise, 117 in tan, 99 in black, 62 in lightcyan) and conducted SPIA. This analysis showed what biological pathways were affected by module genes that were differentially expressed in recurrent MDD. We found that 13 biological pathways were dysregulated in recurrent MDD at FDR ≤ 5%: 12 in turquoise module and 1 in tan module; there were no pathways identified for black and lightcyan modules (Table 3.3). In turquoise module, the following pathways were activated: *Fc gamma R-mediated phagocytosis, Shigellosis, mTOR signaling pathway, Antigen processing and presentation, Herpes simplex infection, Insulin*

signaling pathway, Pathogenic Escherichia coli infection, RNA degradation, Epstein-Barr virus infection, Protein processing in ER, and *Apoptosis* pathways*; while T cell receptor signaling* pathway was inhibited*.* SPIA was also consistent with the over-representation analysis (Table 3.2) and further detailed that *protein processing in ER* pathway was activated in recurrent MDD (FDR adjusted global p-value = 0.04). *Natural killer cell mediated cytotoxicity* was the only affected (inhibited) pathway in tan module.

Table 3.3. Signaling Pathway Impact Analysis (SPIA) on the modules associated with recurrent MDD in older adults.

KEGG pathway name	ID	NP	NDE	Official gene name	pNDE	tA	pPERT	pG	pGFdr	Status
Turquoise										
Fc gamma R-mediated phagocytosis	4666	95	4	*RPS6KB1, ARPC5, WASF2, ASAP1*	0.06	1.00	5.0E-06	1.1E-05	0.001	Activated
Shigellosis	5131	61	8	*ITGB1, WASF2, ARPC5, MAPK9, RIPK2, CHUK, UBE2D2, FBXW11*	3.9E-06	0.27	0.24	0.0001	0.006	Activated
mTOR signaling pathway	4150	64	6	*RRAGB, RRAGC, CAB39, PRKAA1, RHEB, RPS6KB1*	0.0004	0.31	0.04	0.0002	0.006	Activated
Antigen processing and presentation	4612	76	4	*KIR2DL1, KIR2DS5, CTSB, CTSS*	0.03	0.25	0.00	0.0005	0.01	Activated
Herpes simplex infection	5168	184	13	*MAPK9, CHUK, MAP3K7, SKP2, PPP1CB, PPP1CC, IFNGR1, FAS, CASP3, SFRS2B, DAXX, HMGN1, HNRPK*	5.5E-06	0.22	0.43	0.0006	0.01	Activated
Insulin signaling pathway	4910	138	10	*SOCS4, SOCS2, MAPK9, PRKAG2, RHEB, RPS6KB1, PPP1CB, PDE3B, CALM2, PRKAR1A*	5.3E-05	0.63	0.39	0.002	0.03	Activated
Pathogenic Escherichia coli infection	5130	54	4	*TUBB4B, ARPC5, NCK1, ITGB1*	0.009	0.57	0.04	0.002	0.03	Activated
RNA degradation	3018	71	9	*EXOSC3, DCP2, XRN1, PAPD7, SKIV2L2, DHX36, CNOT7, TOB1, HSPD1*	1.3E-06	0.06	0.77	0.003	0.04	Activated
Epstein-Barr virus infection	5169	202	11	*POLR2B, YWHAG, PSMD14, HDAC2, SKP2, CDKN1B, MAP3K14, CHUK, RIPK1, MAP3K7, MAPK9*	0.0003	0.17	0.32	0.003	0.04	Activated
Protein processing in endoplasmic reticulum	4141	166	12	*STT3B, DNAJB11, DNAJC10, TRAM1, DERL1, SSR1, NPLOC4, UBQLN2, MAPK9, UBE2D2, UBE2D3, DERL1*	9.9E-06	0.09	0.67	0.003	0.04	Activated
Apoptosis	4210	87	7	*FAS, RIPK1, CASP3, DFFB, ATM, MAP3K14, CHUK*	5.5E-05	0.26	0.53	0.004	0.04	Activated
T cell receptor signaling pathway	4660	108	7	*NCK1, RASGRP1, MAP3K7, MAPK9, CHUK, MAP3K14, MAP3K8*	0.001	-0.50	0.26	0.005	0.046	Inhibited
Tan										
Natural killer cell mediated cytotoxicity	4650	137	2	*KIR3DL2, TNFRSF10C*	0.001	-1.83	0.02	0.0002	0.008	Inhibited
Black										
Wnt signaling pathway	4310	151	1	*CHD8*	0.15	-0.51	0.08	0.04	0.39	Inhibited
Lightcyan										
Cytokine-cytokine receptor interaction	4060	265	1	*IL24*	0.01	0	1	1	1	Inhibited

NP - the number of genes on the pathway; NDE - the number of DE genes per pathway; tA - the observed total perturbation accumulation in the pathway; pNDE - the probability to observe at least NDE genes on the pathway using a hypergeometric model; pPERT - the probability to observe a total accumulation more extreme than tA only by chance; pG - the p-value obtained by combining pNDE and pPERT; pGFdr - the False Discovery Rate adjusted global p-values; Status - the direction in which the pathway is perturbed (activated or inhibited). A web link to the KEGG website that displays the pathway image with the differentially expressed genes highlighted in red (see Supplementary Materials)

Replication

After QC and pre-processing following the same criteria as in the discovery analyses in sMAS, 11 685 probes for 186 individuals aged 69 and over from the OATS cohort (age range 69.4 – 93.5, mean age 75.9±5.3, 72 males and 114 females) were utilised for replication. High correlations for ranked expression (cor=0.99, $p<e-200$) and connectivity (cor=0.87, $p<1e-200$) and high module preservation calculated for 10 654 probes expressed in both datasets indicated these cohorts were comparable (details in Appendix A).

Significant overlap between MDD-related genes in discovery and replication cohorts

We computed MEs for turquoise, tan, black, and lightcyan modules in OATS and examined their correlations with a binary phenotype of recurrent MDD (N = 7 cases), age, sex, and BMI. Although, there was no association between the MEs and recurrent MDD in the replication dataset (Table S1.6), the directions of associations were consistent with the discovery results (Figure 3.1). Despite the lack of replication at the eigengene level, we observed a significant overlap of 9 individuals genes across four modules ($p<0.03$) that showed association with recurrent MDD in both cohorts ($p<0.05$): 7 genes in turquoise module (*CTSC, ORMDL1, NARG1L, B4GALT4, GTF2H1, AGAP6,* and *THEMIS),* 1 gene in tan module *(IL5RA),* and 1 gene in lightcyan module *(SNX22),* whereas no genes replicated in black module. These 9 genes showed moderate to high correlation with the ME of the relevant module respectively ($0.3>MM<0.9$) of similar strength of correlations across both cohorts. The directionality of dysregulation (GS) was also consistent for 8 out 9 genes across both cohorts (Appendix A, Table S1.7). There was no correlation between ME and age, sex, or BMI, suggesting that our replicated findings are independent of the effect of these potential confounders (Appendix A, Table S1.6).

Discussion

We explored the gene co-expression patterns in peripheral whole blood related to lifetime history of MDD, single episode, current, and recurrent MDD subtypes in older adults. While no significant association between gene expression and lifetime depression, single episode, or current depression was found, interesting results were obtained for recurrent MDD that forms a subgroup of depression with higher severity. Consistent with previous genetic findings that depression is a disorder of multiple genes of small effects, we identified four modules of highly co-expressed genes (a total of > 3000 genes) of which the eigengenes were significantly associated with recurrent MDD. We found that eleven biological pathways, namely *Fc gamma R-mediated phagocytosis, Shigellosis, mTOR signaling pathway, Antigen processing and presentation, Herpes simplex infection, Insulin signaling pathway, Pathogenic Escherichia coli infection, RNA degradation, Epstein-Barr virus infection, Protein processing in ER,* and *Apoptosis* pathways, were activated*;* and two pathways, *T cell receptor signaling* and *Natural killer cell mediated cytotoxicity,* were inhibited in recurrent MDD*.* While we could not replicate association between the eigengenes and recurrent MDD in the independent cohort, which is likely due to insufficient statistical power, there was a significant overlap of recurrent MDD-related individual genes: 9 genes (*CTSC, ORMDL1, NARG1L, B4GALT4, GTF2H1, AGAP6, THEMIS, SNX22, and IL5RA)* were associated with recurrent MDD with highly preserved co-expression and dysregulation patterns between the two cohorts.

There are several notable results that we could not replicate.

Downregulated ribosomal protein (RPS6KB1) in Fc gamma R-mediated phagocytosis and mTOR signalling

The top biological pathway associated with recurrent MDD was *Fc gamma R-mediated phagocytosis,* which plays an essential role in host-defence mechanisms through the

uptake and destruction of infectious pathogens. Activation of this pathway was represented by downregulated *RPS6KB1* and *ARPC5* and upregulated *WASF2* and *ASAP1.* It is worth noting that dysregulation of these genes was consistent in both discovery and replication cohorts, however, we failed to replicate the association of these genes with recurrent MDD at the accepted level of significance. The top gene involved in activation of *Fc gamma R-mediated phagocytosis* was ribosomal protein S6 kinase B1 *(RPS6KB1),* which encodes a protein that responds to mTOR signalling to promote protein synthesis, cell growth, and cell proliferation. *RPS6KB1* was one of the hub genes in turquoise module in both sMAS (MM = 0.83) and OATS (MM = 0.89), indicating its important coordinating role in downstream processing. Consistent with our blood findings, *RPS6KB1* was previously found downregulated in the PFC (Brodmann's area 10) in patients who experienced depression for an average of 9.6 (± 3.6) years and died from suicide (Jernigan et al., 2011). Findings that dysregulation of *RPS6KB1* is detectable in the brain and observable in peripheral blood, provide evidence that *RPS6KB1* may be a potential biomarker for depression. Interestingly, *RPS6KB1* was also recently proposed as a novel antidepressant target; it has been shown that enhanced *RPS6KB1* activity in the medial prefrontal cortex produced antidepressant-like effects and resilience to chronic stress, whereas decreased *RPS6KB1* activity produced pro-depressive behaviour in rats (Dwyer et al., 2015).

Activated pathogen-related pathways

Activation of *Fc gamma R-mediated phagocytosis* pathway along with activation of infectious pathogen-related pathways, such as *Shigellosis, Herpes simplex infection, Pathogenic Escherichia coli infection*, and *Epstein-Barr virus infection* in infectious disease-free recurrently depressed individuals may indirectly indicate a possible link between past pathogenic infection and recurrent MDD. Such a link has been previously

observed in epidemiological research. A large nationwide, prospective cohort study found that any history of hospitalisation for infection increased the risk of mood disorders, including depression, in later life by 62% (Benros et al., 2013). Pathogens are known to be able to influence host response processes, long after a viral load was undetectable (Traylen et al., 2011). Importantly, it has been shown that latent viruses can reactivate their replication in response to stress (Coskun et al., 2010) and ageing (Padgett et al., 1998, Thomasini et al., 2017). Furthermore, *SIRT1*, a longevity-promoting gene, which is known to have an antiviral role (Kim et al., 2016), and was previously associated with depression in a GWAS (rs12415800, p=1.92e-08) (Cai et al., 2015), was also downregulated in recurrent MDD in our study (r = -0.11, p = 0.01). Moreover, *SIRT1* was one of the hub genes in turquoise module (MM = 0.86), suggesting an important regulatory role in depression in older adults. Although speculative, our study suggests that there is a link between activation of pathogen-related pathways and depression in later life.

Dysregulated protein processing pathways in the ER and innate and adaptive immune responses

Emerging evidence suggests that dysregulated protein processing in the ER, leading to accumulation of misfolded proteins and causing ER stress, plays an important role in the pathophysiology of depression (Gold et al., 2013, Gold, 2015). In our study, we found that 11 ER genes were dysregulated in recurrent MDD (*STT3B, DNAJB11, DNAJC10, TRAM1, DERL1, SSR1, NPLOC4, UBQLN2, MAPK9, UBE2D2, UBE2D3)* suggests impaired protein processing in the ER.

We also found that several immune response-related pathways, such as *antigen processing and presentation, T cell receptor signalling, apoptosis, and NK cell-mediated cytotoxicity* are dysregulated in recurrent MDD. It is well known that impaired protein processing in the ER leads to ER stress. To prevent ER stress-related cell damage, the

ER uses unfolded protein response signalling to influence immune responses such as antigen presentation and immunoglobulin synthesis (Janssens et al., 2014). Substantiating this, *antigen processing and presentation* pathway was found to be activated by upregulated *KIR2DL1, KIR2DS5 (*killer cell immunoglobulin-like receptors)*, CTSB* (amyloid precursor protein secretase), and downregulated *CTSS* (involved in the degradation of antigenic proteins to peptides for presentation on MHC class II molecules) in recurrent MDD.

Our findings are in line with the largest transcriptomic study of MDD (Jansen et al., 2016), which identified that *NK cell-mediated cytotoxicity* pathway is implicated in MDD. Specifically, the authors found that *TNFRSF10C* was associated with a change in MDD status over 2-years, i.e. while upregulated in current MDD, the mean expression decreased more in recovered MDD than in the controls. Consistent with these findings, we show that upregulated *TNFRSF10C* and *KIR3DL2* can induce the inhibition of *NK cell-mediated cytotoxicity* pathway in recurrent MDD. We also support recent findings on changes in immunological profile in MDD (Leday et al., 2018). Authors proposed the transcriptional biomarkers panel, consisting of 165 genes differentially expressed between MDD cases and controls. We observed that 24 genes from this panel belong to the recurrent MDD-associated modules in our study (Supplementary materials, Table S15).

The notable strength of this study is the utilisation of various clinical subtypes of depression in conjunction with a longitudinal assessment of participants. However, use of blood, the small size of depression subgroups, reliance on self-report for diagnosis, and a single time point for gene expression are the major limiting factors in our study. Using a population design for our analyses (i.e. each subgroup of MDD was contrasted against the rest of the population irrespective their depression status) can be viewed as a limitation since the control group could include people with a history of depression;

however, this might be of value in that it allows to compare different depression groups against each other (e.g. current vs recurrent vs lifetime). WGCNA as a non-supervised exploratory technique might be useful for this approach as this method does not require prior categorisation of the samples by phenotype, which allowed us increasing robustness of the findings by including all the samples available. In contrast, it is worthwhile to use a control group without history of any form of MDD albeit the resulting smaller sample size. To evaluate this approach using a control group without any form of depression, we conducted additional analyses utilising a reduced sMAS sample (N=465) consisting of recurrent MDD cases (N=27) and healthy controls only (N=438). We observed a large agreement between the results for both types of analyses: four modules were associated with recurrent MDD in a case/control design (Turquoise r=-0.12, p=0.01; Tan r=0.1, 0.04; Black r=0.09, p=0.045; Lightcyan r=0.09, p=0.046) in comparison to a population design (Turquoise r=-0.12, p=0.007; Tan r=0.1, 0.02; Black r=0.09, p=0.04; Lightcyan r=0.09, p=0.04) (Figure 3.1). Another important limitation of this study is that not all potential confounders were examined. Thus, while both our cohorts represent part of a relatively healthy community-dwelling aged Australian population, multiple medical comorbidities, inevitably accompanying ageing processes, and smoking that is known to affect gene expression (Vink et al., 2017, Huan et al., 2016), may have influenced the results. Given the ageing cohorts utilised in our analysis, we cannot rule out the possibility that our findings reflect a combined effect of MDD status and the ageing process itself.

Conclusions

This study supports the epidemiological link between the legacy of infectious diseases and their role in dysregulated biological pathways implicated in later life depression. Replication of our findings and further research utilising a network approach in a larger primary clinical cohort stratified for subtypes of depression is warranted. Specifically,

future research on the long-term effects of interaction between past infectious diseases and depression in later life is recommended.

Supplementary data associated with this article, can be found in the online version, at https://doi.org/10.1016/j.jpsychires.2018.09.017

In the third chapter, the peripheral blood genome-wide gene expression signature of depression using the co-expression network approach was investigated. It was found that over 3,000 highly co-expressed genes were dysregulated in depression, which supports previous findings of depression being a highly heterogeneous disorder associated with multiple genes each with a small effect size. Using WGCNA, we identified these genes as forming four relatively unrelated clusters of highly co-expressed genes within the blood transcriptome network. Collectively, these genes are involved in several molecular pathways, such as protein processing in endoplasmic reticulum and adaptive and innate immune responses, which were found to be dysregulated in the disorder. Although these findings provide a comprehensive view of possible biological mechanisms underlying depression, it failed to replicate network-level findings on an independent cohort (largely due to a limited statistical power). Therefore, a possible translation of these findings into clinical settings is limited. To evaluate whether the mRNA findings in recurrent MDD can be potentially utilised as therapeutic targets or to inform better diagnostic decisions in clinical practice, there is a need to explore whether predictive markers of MDD can be identified when studying complex gene interactions Previous chapters 2 and 3, determined linear associations between mRNA levels and depression. However, in light of the non-replicated findings it is highly unlikely for gene interactions in relation to MDD to be in a linear fashion. Therefore, application of statistical tools which are only able to capture linear interactions are not sufficient to uncover the complexity of molecular dysregulation in depression. To advance the field in this area requires the application of novel methodologies that extend beyond fitting a linear function onto transcriptome data. To this end, we applied machine learning techniques which accounts for both linear and non-linear interactions between mRNA expression levels and depression.

In the next chapter, the predictive capacity of blood transcriptome in recurrent MDD using the same elderly cohort (sMAS) will be evaluated. Using fuzzy forests (FF) approach, a novel machine learning algorithm, we develop a predictive model that assesses the predictive capacity of gene expression levels in recurrent MDD. Fuzzy forests were designed to overcome limitations seen in machine learning literature related to a large feature space relative to sample size in the presence of correlated features, which are known characteristics of transcriptome data that can limit the performance of the classifier. This study will help in identifying the top molecular predictors of recurrent MDD that may potentially serve as predictive markers for the disease.

Statement of Authorship

Title of Paper	Downregulated transferrin receptor in the blood predicts recurrent MDD in the elderly cohort: a fuzzy forests approach
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	CIOBANU, L. G., SACHDEV, P. S., TROLLOR, J. N., REPPERMUND, S., THALAMUTHU, A., MATHER, K. A., COHEN-WOODS, S., STACEY, D., TOBEN, C., SCHUBERT, K. O. & BAUNE, B. T. 2019. Downregulated transferrin receptor in the blood predicts recurrent MDD in the elderly cohort: a fuzzy forests approach. *Journal of Psychiatric Research* (under review).

Principal Author

Name of Principal Author (Candidate)	CIOBANU, L. G.		
Contribution to the Paper	Conceived of the presented research concept and design , analysed clinical data to derive recurrent depression subtype, generated and pre-processed gene expression data, performed machine learning analysis, interpreted the results, wrote the manuscript and submitted to the journal		
Overall percentage (%)	85%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	20.05.2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	SACHDEV, P. S.		
Contribution to the Paper	• sMAS cohort study design and data collection • Critical revision of the manuscript • Discussed the results and contributed to the final manuscript		
Signature		Date	11 June 2019
Name of Co-Author	TROLLOR, J. N.		
Contribution to the Paper	• sMAS cohort study design and data collection • Critical revision of the manuscript • Discussed the results and contributed to the final manuscript		
Signature		Date	11/06/2019
Name of Co-Author	REPPERMUND, S.		
Contribution to the Paper	• sMAS cohort study design and data collection • Critical revision of the manuscript		
Signature		Date	11/06/2019

Name of Co-Author	THALAMUTHU, A.		
Contribution to the Paper	• sMAS cohort study design and data collection • Critical revision of the manuscript • Discussed the results and contributed to the final manuscript		
Signature		Date	12/06/2019
Name of Co-Author	MATHER, K. A.		
Contribution to the Paper	• sMAS cohort study design and data collection • Critical revision of the manuscript		
Signature		Date	6/6/19
Name of Co-Author	COHEN-WOODS, S.		
Contribution to the Paper	• Critical revision of the manuscript		
Signature		Date	29/05/2019
Name of Co-Author	STACEY, D.		
Contribution to the Paper	• Verified the analytical methods • Critical revision of the manuscript		
Signature		Date	31/05/2019
Name of Co-Author	TOBEN, C.		
Contribution to the Paper	• Critical revision of the manuscript • Helped shape the final manuscript		
Signature		Date	10 05 2019
Name of Co-Author	SCHUBERT, K. O.		
Contribution to the Paper	• Critical revision of the manuscript		
Signature		Date	21. 5. 19
Name of Co-Author	BAUNE, B. T.		
Contribution to the Paper	• Supervised the project • Provided critical feedback to study design and methodology • Provided critical feedback to the first draft of the manuscript • Discussed the results and contributed to the final manuscript • Final decision on manuscript submission		
Signature		Date	16.06. 2019

CHAPTER 4

Downregulated transferrin receptor in the blood predicts recurrent MDD in an elderly cohort: a fuzzy forests approach

Liliana G. Ciobanu[1,2], Perminder S. Sachdev[3], Julian N.Trollor[3,4], Simone Reppermund[3,4], Anbupalam Thalamuthu[3], Karen A. Mather[3,7], Sarah Cohen-Woods[5], David Stacey[6], Catherine Toben[1], K. Oliver Schubert[1,8], Bernhard T. Baune[9,10,11]*

[1]Discipline of Psychiatry, Adelaide Medical School, The University of Adelaide, South Australia, Australia
[2]Quality Use of Medicines and Pharmacy Research Centre, School of Pharmacy and Medical Sciences, University of South Australia, Australia
[3]Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, UNSW Sydney, New South Wales, Australia
[4]Department of Developmental Disability Neuropsychiatry, School of Psychiatry, UNSW Sydney, New South Wales, Australia
[5]School of Psychology, Flinders University, South Australia, Australia
[6]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
[7]Neuroscience Research Australia, Randwick, Australia
[8]Northern Adelaide Local Health Network, Mental Health Services, Lyell McEwin Hospital, Elizabeth Vale, South Australia, Australia
[9]Department of Psychiatry and Psychotherapy, University of Münster, Münster, Germany
[10]Department of Psychiatry, Melbourne Medical School, The University of Melbourne, Melbourne, Australia
[11]The Florey Institute of Neuroscience and Mental Health, The University of Melbourne Parkville, VIC, Australia

ABSTRACT

At present, no predictive markers for Major Depressive Disorder (MDD) exist, and there is limited understanding of the biological underpinnings of depression. The search for such markers is challenging due to clinical and molecular heterogeneity of MDD, the lack of statistical power in studies and suboptimal statistical tools applied to multidimensional

data. Machine learning is a powerful data mining approach that has been proven successful in biomedical research. Transcriptomic data are highly multidimensional with the presence of correlated features organised in a network-like structure. To develop a meaningful predictive model using this type of data, a large number of observations are required; however the typical transcriptomic dataset is of a relatively small sample size. Multidimensional space with correlated features, coupled with a lack of observations, posits a challenge for machine learning classification. In an attempt to overcome these limitations, we adopted a novel Fuzzy Forests approach that takes advantage of the co-expression network structure between genes aiming to identify predictive markers for recurrent MDD in the elderly at the blood gene expression level. Utilising transcriptome data from 521 individuals in the Sydney Memory and Aging Study (sMAS), we developed a model that correctly classified 63% of recurrently depressed individuals in the test dataset. We found that the most predictive marker, the gene *TFRC*, which encodes transferrin receptor, is downregulated in recurrent MDD and may represent a predictive marker for recurrent MDD.

INTRODUCTION

Currently investigations into biological underpinnings of MDD remain challenging; however it is paramount for developing reliable diagnostic tools and effective treatments. Despite decades of research, elucidation of the exact molecular mechanisms is in its infancy (Cai et al., 2015, Okbay et al., 2016, Hek et al., 2013, Wray et al., 2018, Hyde et al., 2016, Jansen et al., 2016). MDD as a heterogeneous disorder is a complex dynamic system from both clinical (Cramer et al., 2016) and biological (Sibille and French, 2013) perspectives. The biological complexity of MDD can be accounted for by studying altered gene expression patterns in affected individuals compared to unaffected. These dysregulated patterns can serve as a dynamic marker of the disorder.

As far as molecular biology is concerned, genes do not act in isolation; instead, they interact within each other akin to complex networks that might be disrupted in depression. In our previous study, we explored what genome-wide gene co-expression patterns are associated with depression. We applied Weighted Gene Co-expression Network Analysis (WGCNA) to transcriptomic data from 521 community-dwelling individuals aged over 65. We studied four subtypes of depression derived from the data collected at four time points with 2-year intervals between assessments: single episode, current MDD, recurrent MDD, and lifetime MDD. Although we were not able to detect a signal for the single episode, current or lifetime MDD, we found that four clusters containing 1,241 of highly interacting genes are associated with recurrent MDD status. Using in-silico Enrichment and Signalling Pathway Impact Analysis (SPIA) we found that this gene pool was biologically meaningful with 13 known molecular pathways significantly dysregulated in recurrent MDD in the elderly (Ciobanu et al., 2018b). While these findings were consistent with previous observations, and provided new insights into aetiology of depression, we sought to further explore the complex relationship between recurrent MDD, as the more severe subtype of depression, and transcriptome alterations.

The typical biostatistical approach is to fit linear function between variables and the outcome. Although this approach is powerful in many scenarios, it can be suboptimal for multivariate transcriptome data. Machine learning (ML) provides an alternative view for data analysis, allowing for complex linear and non-linear interactions between the features to be explored. ML explicitly focuses on learning statistical functions from multidimensional data sets to make generalizable predictions about affected individuals.

Random forests (RF) is an established technique for classification and feature selection, owing to its unique advantages in dealing with relatively small sample size, high-dimensional feature space, and complex data structures. While RF is able to capture the true importance of features in settings where the features are independent, it is

established that RF is biased when features are correlated with one another and the correlation structure is not known a priori (Nicodemus and Malley, 2009), which is a typical scenario for transcriptome data. Fuzzy forests (FF) an extension of a RF algorithm, is designed to reduce this bias. FF is a novel algorithm which takes advantage of the network structure between features and relies on WGCNA to create relatively uncorrelated clusters of highly correlated features (Zhang and Horvath, 2005). FF uses recursive feature elimination RF to select features from separate clusters (Díaz-Uriarte and Alvarez de Andrés, 2006). The final RF is fit using the surviving features. The selected features are then used to construct a predictive model (Conn et al., 2015, Conn et al., 2016).

Although FF is based on WGCNA, these methodologies represent two different analytical strategies. WGCNA is primarily concerned with identifying important genes assumed to be involved in the same biological processes, which is useful in understanding biological underpinnings of depression. However, given that depression is a biologically multifactorial disorder, it is likely that hundreds to thousands of genes are involved in the disease, making it diagnostically impractical. RF aims to find a small number of genes sufficient for a good prediction of the response variable. While WGCNA is based on fitting a linear function into data, RF is non-parametric and non-linear. Athough this makes it difficult for interpretation it is more useful for identifying prediction markers and thereby clinical translation. Combining the two strategies in FF framework may help to overcome limitations of each individual method and enrich our understanding of the aetiology of depression.

METHODS

Sample characteristics

The Sydney Memory and Aging Study (sMAS) was initiated in 2005 to examine the clinical

characteristics and prevalence of mild cognitive impairment and related syndromes, including depression, in a non-demented population aged 70-90 years at recruitment (N=1,037) (Sachdev et al., 2010). The phenotypic data were collected at four time points with 2-year intervals between assessments. Blood samples for gene expression analyses were collected at Wave 4 (N=521), six years after baseline data collection.

MDD was assessed by two well-validated self-report, and two clinical interview-based, measures of depression including the Geriatric Depression Scale (GDS-15) (cut-off ≥ 6) (Yesavage et al., 1982a), the Patient Health Questionnaire (PHQ-9) (cut-off ≥ 10) (Kroenke K and R., 2002), the Neuropsychiatric Inventory (NPI) (depression sub-scale) (Cummings et al., 1994a), and the Mini International Neuropsychiatric Interview (M.I.N.I.) (Sheehan et al., 1998). MDD cases were defined as participants who were identified as clinically depressed by at least two of the above depression measures at any one study Wave, according to DSM-IV criteria. A recurrent MDD subgroup consisted of 27 individuals as published in (Ciobanu et al., 2018b).

Gene expression data pre-processing

Whole blood gene expression data for 521 participants were generated from PAXgene tubes using the Illumina Whole-Genome Gene Expression Direct Hybridization Assay System HumanHT-12 v4 (Illumina Inc., San Diego, CA, USA) according to standard manufacturer protocols. After rigorous QC, filtering, and adjustments for known (RINs, age, sex) and latent non-biological variables, gene expression data for 11,018 probes were available for downstream analyses (Ciobanu et al., 2018b).

Data partitioning and balancing

Our data represent a population study design with extremely unbalanced classes of individuals with recurrent MDD and those without. Since the aim of any classification

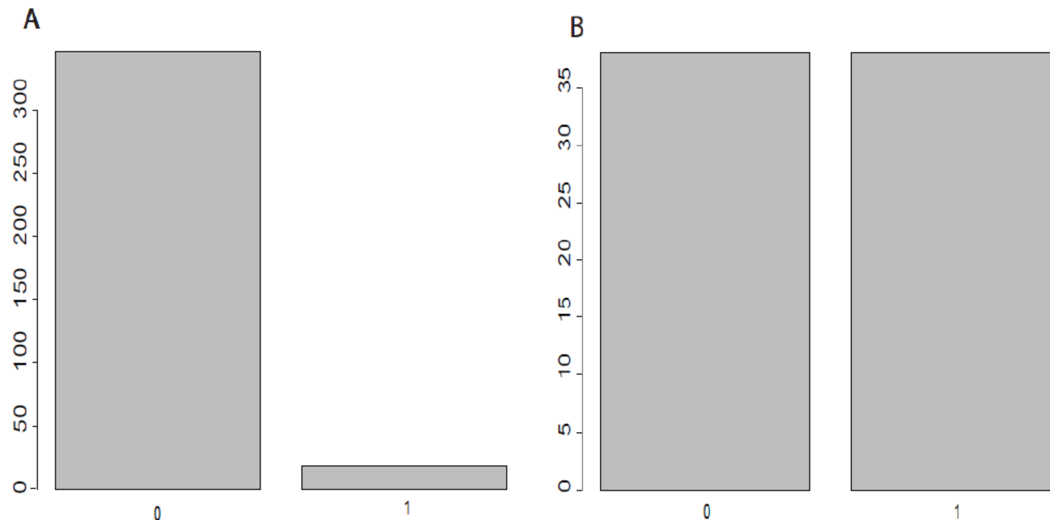
algorithm is to minimize the overall error rate, it posits a challenge for ML algorithms to correctly predict the minority class. Our solution was to train the model on balanced dataset, containing an equal proportion of positive and negative instances, and test it on unbalanced dataset that better reflects population distribution. To balance our training set (70% of the whole dataset), we used Synthetic Minority Over-sampling Technique, SMOTE, a hybrid method allowing for down-sampling the majority class and synthesizing new data points in the minority class using *k*-nearest algorithm, where *k* was set to 5 (Chawla et al., 2002). The never seen before training set (30% of the whole dataset) was left in its original distribution.

RESULTS

Training and test data

After partitioning the full dataset, our training consisted of two groups: 19 recurrently depressed individuals (group [1]) and 346 individuals without recurrent MDD (group [0]) (0.05 vs 0.95), which is highly unbalanced. Using SMOTE, a combined method of over-sampling the minority (MDD) class and under-sampling the majority (non-MDD) class, we balanced training data to 38 observations in each group (Figure 4.1). The test data consisted of 8 recurrent MDD [1] and 148 non-recurrent MDD [0] individuals.

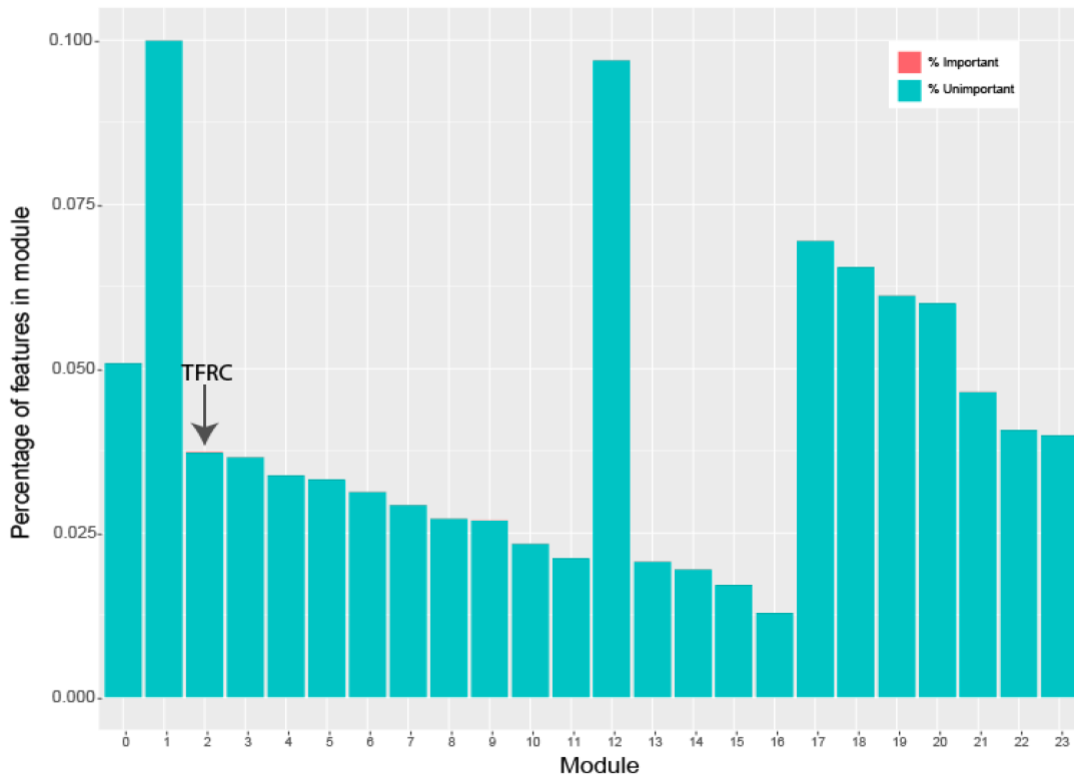
Figure 4.1. Balancing training dataset using SMOTE. A. Unbalanced dataset; B. Balanced dataset.



Modules of co-expressed genes and FF model performance

To determine the power of the adjacency function, we estimated network topology for various soft-thresholding powers. Based on these analysis, we used the scale-free topology criterion β = 6 (more details in Supplementary materials, Figure S1). Minimum module size, *minModuleSize*, was set to 100. Using WGCNA within the FF, we constructed a co-expression network consisting of 23 modules of co-expressed genes. Following the FF workflow, we applied recursive feature elimination RF to select features from separate modules. The last RF was performed on the genes from the survivors list. After parameter tuning, the best performance was achieved using the following hyperparameters: *min_ntree* = 500, *final_ntree* = 500, *mtry factor* = 5, *ntry_factor* = 5, *drop_fraction* = 0.25, *keep_fraction* = 0.05, *number_selected* = 1 (details in the Appendix A). This final model was tested on the test dataset achieving sensitivity of 63%, specificity of 66% and balanced accuracy of 63% for the top predictive feature – gene *TFRC* (ILMN _1674243) assigned to the module 2. (Figure 4.2).

Figure 4.2. Module membership distribution plot visualizing the relative importance of the modules for recurrent MDD. The blue bars represent the percentage of selected features in a particular module. The top predictive feature of recurrent MDD, gene *TFRC*, was assigned to the module 2.



TFRC is implicated in recurrent MDD – consistent findings from co-expression network analysis, fuzzy forests machine learning algorithm and differential expression analysis.

As described above, FF approach captures both non-linear and linear interactions between the genes to predict the outcome. Although we cannot directly estimate how much each type of interaction contributes to the final model, we sought to determine the effect of linear relationship between *TFRC* levels and recurrent MDD using differential expression analysis for the identified candidate gene, and, therefore, to establish the directionality of dysregulation of *TFRC* in recurrent MDD in our data. In our previous study, we investigated patterns of transcriptome dysregulations in recurrent depression using linear modelling implemented in co-expression network analysis (Ciobanu et al., 2018b) and found that *TFRC* was assigned to a cluster of co-expressed genes, the

eigengene of which was associated with recurrent MDD (p=0.007). At the individual gene level, there was a significant difference of a small effect (at the FDR 20%) in the *TFRC* expression in recurrent MDD compared to those without this diagnosis in the sMAS cohort (logFC = -0.14), suggesting that *TFRC* is downregulated in people diagnosed with recurrent MDD. Although we could not replicate this finding in the independent replication cohort the Older Australian Twins Study (OATS) at the accepted level of significance, which was largely due to the limited sample size, it is worth noting that directionality of dysregulation was consistent with the primary findings (logFC = -0.01) (details in Ciobanu et al. (2018b), Appendix A).

DISCUSSION

Downregulated transferrin receptor, TFRC, as a potential predictive marker for recurrent MDD

In this study, we applied a novel ML algorithm on blood transcriptome data in order to identify a biological predictive marker in recurrent MDD. Using FF we developed the model that was able to correctly classify 63% of recurrently depressed elderly individuals on a test dataset. We found that the transferrin receptor gene, *TFRC,* was the top predictor of recurrent MDD in our cohort, suggesting that *TFRC* can potentially serve as a predictive marker for recurrent MDD. This finding was consistent with our previous study using this cohort, where downregulated *TFRC* was assigned to the module of interconnected genes the eigengene of which was associated with recurrent MDD (Ciobanu et al., 2018b). Taken together, we observed that three separate methodologies, namely WGCNA, FF, and differential expression produced comparative results, pointing towards *TFRC* as being involved in recurrent MDD. *TFRC* was also previously found dysregulated in the brains of MDD patients suffering from disrupted circadian rhythm (Li et al., 2013).However, it is unlikely that *TFRC* is the only gene that has a predictive

capacity given the molecular complexity of depression. Therefore, our model cannot be considered comprehensive. Instead a model capturing a larger pool of genes is likely to have a better predictive performance which could be achieved with a larger sample size. This will ensure the model is trained on sufficient number of observations to learn a true pattern of interactions between multiple genes.

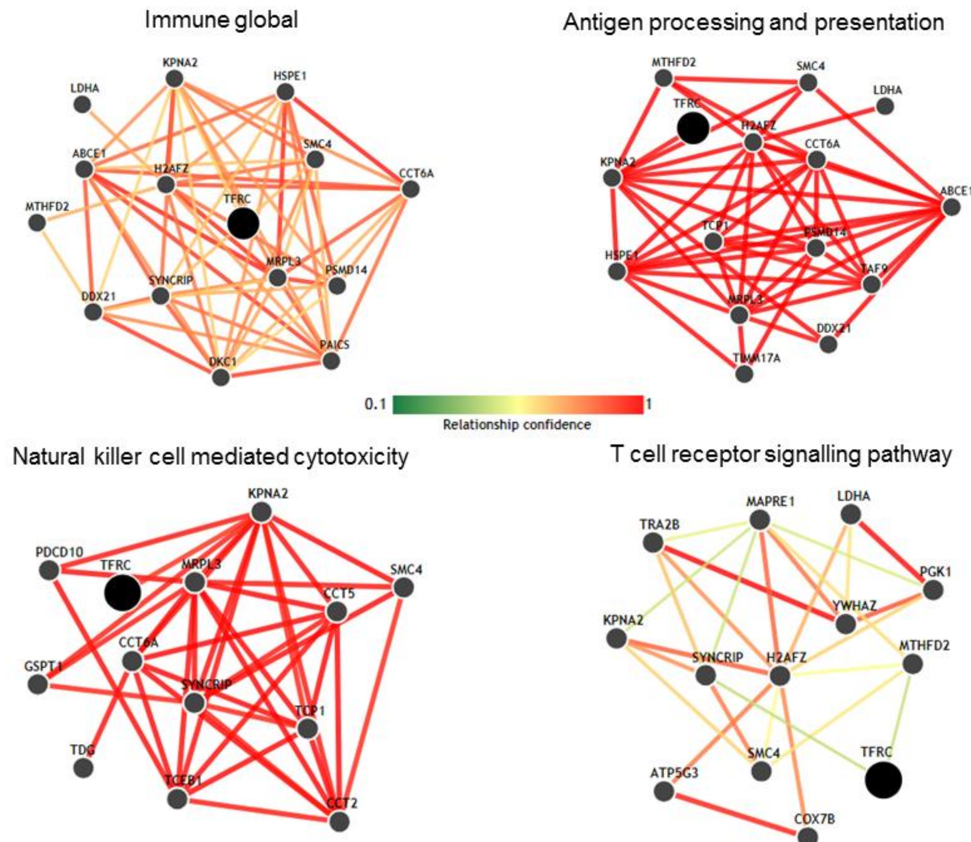
Transferrin receptor gene, TFRC, the gatekeeper of cellular iron uptake

Transferrin receptor is a type II transmembrane glycoprotein expressed by all nucleated cells of the body. *TFRC* is an important modulator of iron homeostasis recognised as the gatekeeper of cellular iron uptake (Porto and De Sousa, 2007). Expression of *TFRC* is regulated at both the transcriptional and post-transcriptional levels by the cellular iron status: in a cellular iron-deficient state, *TFRC* expression increases, whereas in the presence of excess iron, *TFRC* expression decreases (Walker and Walker, 2000, Khumalo et al., 1998, Kawabata, 2018). Downregulated mRNA *TFRC* in blood, therefore, may indicate a disrupted iron homeostasis in depression (Brandao et al., 2005, Rostoker et al., 2015).

The role of transferrin receptor in immunity

Transferrin receptor plays a role in many immune-related processes, such as *antigen processing and presentation, natural killer cell mediated cytotoxicity* and *T cell receptor signalling* (Figure 4.3); these pathways were also previously found disrupted in depression (Haapakoski et al., 2016).

Figure 4.3. The *TFRC* immune-related functional relationships. The networks were constructed using ImmuNet software (Gorenshteyn et al., 2015).



In our previous study, we found that multiple genes were associated with dysregulated immune system-related processes and activation of four pathogen-related pathways (*Shigellosis, Herpes simplex infection, Pathogenic Escherichia coli infection,* and *Epstein-Barr virus infection*) in recurrent MDD, providing support for the role of previously observed complex interaction between pathogens and host's immune response in depression (Carter, 2013, Ciobanu et al., 2018b). Furthermore, an emerging body of evidence indicates that many viruses and parasites use the transferrin receptor, for cellular entry into host cells (Weinberg, 1996). Downregulated *TFRC* coupled with dysregulated immune-related pathways therefore may indicate interplay between altered pathogen and host response mechanisms in recurrent MDD. Although we report on the

ability of our model to predict recurrently depressed individuals, these results should be treated with caution. While we identified *TFRC* as the most predictive gene for recurrent depression in the elderly our test sample was fairly small, relative to the feature space and could be a source of poor generalizability. While we corrected our data for age, sex, RINs and latent non-biological variables, we were unable to account for medications taken, comorbidities, and other environmental confounders that can be numerous in the elderly cohort. Furthermore, the serum levels of transferrin receptor were not assessed in this study.

Using fuzzy forests framework, we identified that the most predictive gene, *TFRC*, can predict recurrent depression in the elderly with an accuracy of 63%. This finding, coupled with our previous observation that blood *TFRC* mRNA downregulated in recurrent MDD individuals as compared with those without, may potentially serve as a recurrent MDD-specific predictive marker and provide some insights into pathophysiology of depression. Although our study is exploratory in nature providing preliminary results which require further exploration, the use of ML in biological psychiatry is an emerging field which will promote clinical translation and inform personalised psychiatry in the future.

An Emerging Need to Focus on Immune Dysregulation in Depression

In the previous chapters, depression-relevant alterations of transcriptome using various approaches were explored, including traditional statistical approaches based on fitting a linear function into data, such as differential expression analysis (Chapter 2) and WGCNA (Chapter 3). Furthermore the machine learning approach enabled predictions to take into account linear as well as non-linear interactions between the genes and phenotype of interest (Chapter 4). While we did not find an absolute agreement between the findings, a common theme has emerged. Immune dysregulation was a highlight in all three studies,

suggesting that focusing on this aspect of transcriptome dysregulation might provide us with further insights into pathophysiology of depression.

In Chapter 5, the research focus is on immune related biological pathways that we previously found dysregulated in recurrent depression. First, a comprehensive overview of previous findings on immune alterations in depression at gene expression level for both brain and periphery is presented. Next, in the experimental part of this chapter, the differential gene expression of immune genes that are known to be involved in immune-related pathways in recurrent MDD on an independent cohort are evaluated.

CHAPTER 5

Dysregulation of immune-related pathways in depression

ABSTRACT

Immune dysregulation is one of the most consistent findings in the biology of depression research. While many 'immune' genes were previously found implicated in depression, including upregulated pro-inflammatory cytokines *IL1B*, *IL6*, *TNF*, and *INF,* and transcription factors *FN-kB* and *CREB1* identified in both the brain and the periphery, the results are inconsistent across studies. Furthermore, at the transcriptome level, dysregulation of immune-related pathways at the periphery beyond the commonly studied cytokines was consistently found in multiple studies including our own. However, the exact mechanisms of peripheral immune dysregulation in depression are still largely unknown. In this study, we (1) provide a comprehensive overview of immune dysregulation in depression at gene expression level in both brain and periphery, and (2) using our previous exploratory findings on 13 immune-related pathways being dysregulated in recurrent MDD (Chapter 3), we examine the role of these pathways in recurrent MDD at individual gene level on independent cohort (the Older Australians Twin Study, OATS). To target the immune pathways, we selected all known genes (KEGG) involved in these 13 pathways (N = 1,302) and conducted a differential expression analysis on these candidates between individuals with recurrent MDD (N = 186). We found that two Illumina probes corresponding to *CD14* were significantly downregulated in recurrent MDD (FDR < 0.05). This suggests that *CD14* can potentially serve as a peripheral marker of immune dysregulation in recurrent MDD.

Gene expression of inflammation in depression presents as a relatively novel and promising approach to uncover the pathophysiology of depression and to possibly provide useful clinical information for predicting treatment response and for decision-making processes in depression treatment. Quantifying the abundance of mRNA molecules known to be involved in immune-related pathways in a single cell or in a population of cells provides essential information on the biological activity and functions of immune system. Studying gene expression alteration in genes that are involved in immune-related biological pathways in depression can be viewed as complementary to a gene discovery approach, both aiming at understanding the molecular mechanisms of depression. There is a body of research that investigates gene expression alterations in depression in the brain as well as in the periphery at genome-wide level (when all genes are assessed) as well as using gene targeted (when several candidate genes are assessed) approaches. These studies have substantially improved our understanding the biology of depression; in particular, it is now clear that the immune system plays a role in its aetiology. However, to the best of our knowledge, there were no depression-related pathway-targeted studies aiming to assess all genes known to be involved in a particular molecular pathway, which is essential to better understanding the role of the immune system in depression. In this study, we aim to answer two questions: (1) what genes known to be involved in specific pathways are dysregulated in depression and (2) do these genes perform in isolation in one pathway, or have a pleiotropic effect on multiple immune pathways? First, we provide a comprehensive overview of known alterations in expression of immune-related genes in the brain and in the periphery (this overview was published as a part of a book chapter (Ciobanu and Baune, 2018), the full chapter can be accessed via https://doi.org/10.1016/B978-0-12-811073-7.00011-8), and secondly, using a pathway-

targeted approach, we examine the expression levels of all genes known to be involved in 13 immune-related pathways that were found associated with depression.

Gene expression patterns of immune dysregulations in the brain

Studying the gene expression patterns in post-mortem brain tissues of individuals who suffered from depression provides us with valuable information about molecular changes occurring in depressed brains compared to healthy controls. Such studies have substantially advanced our understanding of the pathophysiological mechanisms of depression. Gene expression signatures derived from various brain regions collectively point towards molecular processes involving inflammatory pathways, cell survival, apoptosis and oxidative stress (Bakunina et al., 2015).

Structural and functional neuroimaging studies in humans suggest that the limbic system (predominantly, amygdala (Ciobanu and Baune, 2018) and hippocampus) and the prefrontal cortex serve as primary brain areas responsible for disturbances in emotion processing and mood regulation in depression (Wise et al., 2014). Gene expression studies on human brain tissue, utilizing both candidate genes and genome-wide approaches, provide some support for dysregulated immune signalling within the brain, however the results lack consistency across the studies, which makes it challenging to specify how altered markers of inflammation found in different brain areas contribute to depression. For instance, the most commonly reported circulating markers of inflammation, pro-inflammatory cytokines *IL1B, IL6, TNF* or *INF* were upregulated within various areas of prefrontal cortex (PFC), such as the dorsolateral prefrontal cortex, DLPFC (BA9) (Kang et al., 2007) and the anterior PFC (BA10) (Shelton et al., 2011, Malki et al., 2015) of depressed individuals, in the primary ventral regions of the prefrontal cortex (BA 44, 45, 46, 47) of depressed suicide victims (Klempan et al., 2009), in the orbitofrontal area (BA11) of adult suicide victims (Tonelli et al., 2008), in BA8 (part of the

frontal cortex involved in the management of uncertainty) and the anterior PFC (BA10) of teenage suicide victims (Pandey et al., 2012). However, none of these genes were replicated within the same brain area. Immune and apoptosis signalling along with synaptic and glutamatergic signalling pathways were also found disrupted in the hippocampal subfields DG and CA1 of middle-aged subjects diagnosed with MDD (Duric et al., 2013). An interesting study proposing synchronised dysregulation of expression in depression across different brain areas found a shift in coordinated gene expression levels between the amygdala and cingulate cortex for 100 to 250 individual genes, including *IL1* and *CREB1* in male MDD patients (Gaiteri et al., 2010b). Furthermore, several transcription factors known to be involved in immune response were found dysregulated in the depressed brain. For example, alterations in expression levels of *CREB1,* a transcriptional factor known to be involved in a wide variety of biological processes including immune response, is one of the most consistently replicated findings, however, the directionality of dysregulation is not consistent across different brain areas. For instance, Sibille et al. (2004) found that *CREB1* is downregulated in the DLPFC (BA9 and BA47) of depressed suicide subjects (noting that it did not survive correction for multiple testing), while Tochigi et al. (2008) observed upregulation of *CREB1* in the anterior PFC (BA10) of non-suicide depressed subjects. This discrepancy may be explained by the presence of a suicide component in one study and its absence in another, pointing towards a differential role of *CREB1* in depression and suicide. Alternatively, it may indicate that *CREB1* is downregulated in the DLPFC and upregulated in the anterior PFC in depression. Another transcription factor *FOXD3*, which functions as a transcriptional repressor, was found upregulated in the DLPFC together with *TNFRSF11B, INFA6,* and *INFR1* (Kang et al., 2007). Moreover, it seems that post-transcriptional regulation by short non-coding microRNAs, affecting both the stability and translation of mRNAs, are also involved in depression. Thus, utilising RNA-sequencing

data derived from DG granule cells, it has been found that posttranscriptional regulation by *miR-182*, which is involved in a broad range of biological processes including regulation of immune response, significantly contributed to disrupted signalling in the hippocampus (DG) in depression (Kohen et al., 2014b) (Table 5.1).

Although all these findings suggest that brain dysregulation of immune genes, involving *IL1B, IL6, TNF, INF, CREB1, FOXD3,* and *miR-182* might play a role in depression, they do not provide solid evidence, as the majority of those individual players of immune signalling have not been replicated. To further explore the level of replication of the findings in gene-expression studies, we re-analysed data from 15 brain transcriptomic studies in depression and identified only seven genes of the immune response showing minimal replication (i.e. they were found dysregulated within the depressed brain but in different brain areas): *FGFR3, ENPP2, PTP4A2* (innate immune response), *CREB1, MOG*, *JUN* (toll-like receptor signalling pathways), and *LEPR*, which belongs to the gp130 family of cytokine receptors (cytokine-mediated signalling (Ciobanu et al., 2016). These findings point to a possible involvement of immune factors going beyond the typically reported pro-inflammatory cytokines.

Table 5.1. Studies on the dysregulation of immune genes and oxidative stress factors in brain tissues in depression

Citation	Brain Area/Cell Type	Platform/ Genes Studied	Diagnostic Criteria	Sample Size	Medications	Main Findings
Klempan et al. (2009)	PFC: BA44, 45, 46, 47	Affimetrix HG-U133	DSM-IV	16 Depressed suicide and 8 nondepressed suicide cases 13 Controls	One subject in depressed suicide group known to take antidepressants	"Cytokinesis" and "immune cell activation" are the central GO terms for distinguishing between depressed suicides and controls
Kohen et al. (2014)	Hippocampus DG granule cells by LCM	RNA-seq	DSM-IV	Cases (age range 25–91): 17 MDD 17 Schizophrenia 16 Bipolar disorder 29 Controls	NA	Disrupted signaling by miR-182 (loss of DG miR-182 signaling) in SZ and MDD miR182 is involved in a broad range of biological processes, including immune response
Malki et al. (2015)	PFC: BA10	Affimetrix HG-U95A	DSM-IV	11 MDD cases (age <65, M:F) 15 Controls	Antidepressant use included as covariate in analysis	80% of dysregulated genes were functionally associated with of a key stress response signaling cascades, including *NF-kB*, *AP-1*, and *ERK/MAPK*
Shelton et al. (2011)	PFC: BA10	Affimetrix Human Exon 1.0 ST	DSM-IV 11 out of 14— melancholic subtype	14 MDD cases (age mean 47.2±14.0, 11M:3F) 14 Controls	Psychotropic drug-free	Increased inflammatory and apoptotic stress, including elevated cytokines
Sibille et al. (2004)	PFC: BA9, BA47	Affimetrix U133A	DSM-III-R SCID-1	19 Depressed suicide cases (mean age 44.6±21.2 75%M) 19 Controls	Psychotropic medication-free	*CREB1* was downregulated in suicide victims, however, did not reach the level of significance

Continued

Citation	Brain Area/Cell Type	Platform/ Genes Studied	Diagnostic Criteria	Sample Size	Medications	Main Findings
Szebeni et al. (2014)	Astrocytes and oligodendrocytes from temporal lobe (UF) or right BA10 by LCM	*SOD1*, *SOD2*, *GPX*, *CAT*	DSM-IV	12 MDD cases (mean age 51, M:F) 12 Controls	Information about toxicology is available, included as covariate in analysis	Oxidative defense enzymes *SOD1* and *SOD2*, catalase (*CAT*) and glutathione peroxidase (*GPX1*) lower in oligodendrocytes from MDD
Tonelli et al. (2008)	Orbitofrontal cortex: BA11	*TNF*, *IL1B*, *IL4*, *IL5*, *IL6*, *IL16*	Limited information about diagnosis	34 Suicide cases (9 cases of MDD) (mean age 52.3±15, 20M:14F) 17 Controls	Limited information about toxicology	Elevated level of cytokines in suicide victims (*IL4* in females, *IL13* in males)
Tochigi et al. (2008)	PFC: BA10	Affimetrix HU95Av2	DSM-IV	11 MDD cases (nonsuicide) (mean age 46±10, 6M:5F) 11 Bipolar disorder cases 13 Schizophrenia cases 15 Controls	Information about antipsychotic drugs is available	*CREB1* was upregulated in BA10 of nonsuicide MDD subjects

HEEBO, human exonic evidence-based oligonucleotide; *LCM*, laser capture microdissection; *PFC*, prefrontal cortex; *DLPFC*, dorsolateral prefrontal cortex; *PMI*, port-mortem interval; *BA*, Brodmann's area; *UF*, uncinate fasciculus; *DG*, dentate gyrus; *DSM-IV*, Diagnostic and Statistical Manual of Mental Disorders, version IV; *SCID-1*, Structural Clinical Interview for DSM.

It has been suggested that inflammation is tightly linked with oxidative stress in depression, which may lead to cell death and further inflammation, creating a vicious circle, and a mechanism which is not well understood (Bakunina et al., 2015). A recent integrative brain analysis of rat and human prefrontal cortex transcriptomes revealed that 80% of dysregulated genes were functionally associated with a key stress response signalling cascade, involving *NF-kB*, *AP-1* (activator protein 1) and *ERK/MAPK*, suggesting inflammation-mediated oxidative stress and further dysregulation of neuroplasticity and neurogenesis in the prefrontal cortex in MDD (Malki et al., 2015). Furthermore, oxidative stress, measured by expression levels of four antioxidant oxidative defence enzymes, A+B) superoxide dismutases (*SOD1* and *SOD2*), C) catalase (*CAT*) and D) glutathione peroxidase (*GPX1*), which were significantly lower in depressed individuals, found to contribute to telomere shortening in oligodendrocytes of the

prefrontal cortex (BA10) (Szebeni et al., 2014). These findings provide evidence for a possible aetiological link between inflammation, oxidative stress, telomere shortening and white matter abnormalities previously observed in depression.

Although many individual players of the immune and the oxidative stress pathways have been found altered in brain tissue of patients diagnosed with depression, an inconsistency on the direction and brain areas of dysregulated genes precludes firm conclusions on specific immune-related pathways dysregulated in the brain in depression. The current state of knowledge suggests that multiple brain areas are possibly involved in depression-related immune dysregulation in a complex manner.

Overlap of gene expression markers of inflammation across psychiatric disorders

Given that a transdiagnostic assessment may reveal common biological features between depression, schizophrenia and bipolar disorder, the comparative assessment of gene expression levels can shed some light on shared and distinct pathophysiological mechanisms of these disorders, which can potentially provide a molecular basis for developing diagnostic guidelines. Molecular comparisons of different brain areas in schizophrenia, bipolar disorder and MDD at a transcriptome level suggested that (1) the prefrontal cortex in all three disorders have distinct gene expression signatures with shared upregulation of genes encoding proteins for transcription and translation (Iwamoto et al., 2004) and that (2) the activation of immune/inflammatory response in the hippocampus is present in all disorders. However, despite these similarities across disorders, several differences in the specifics of the dysregulated transcriptomes were found. For MDD, abnormal activation of the first component of the complement cascade C1q (hub genes in co-expression analysis *C1QA, C1QB* and *C1QC*) and *IL1B* (Kim et al., 2016) were observed in the hippocampus, while a dysregulation of the immune-related response in the thalamus, including B cell receptor signalling, was specific to

schizophrenia, as it was not found in depression (Chu et al., 2009) (Table 5.1).

Taken together, there is a plethora of findings pointing towards inflammation and oxidative stress-related events in depressed brains, including upregulation of pro-inflammatory cytokines, transcriptional and post-transcriptional regulators of immune signalling in the prefrontal cortex and hippocampus. Emerging evidence also suggests a possible causal link between inflammation, oxidative stress and structural changes in the brain in depression. However, it is premature to propose a distinct inflammatory/oxidative stress transcriptomic signature of depression as of yet, as replication of these findings is minimal at present. Further exploration of transcriptomes across different brain areas at a single cell type level and peripheral blood has a great potential to discover how inflammation-induced molecular changes lead to structural abnormalities and impair neural circuits involved in emotional and cognitive processing in depression.

Peripheral Gene expression patterns of immune dysregulation

Studying gene expression markers of depression in peripheral tissues is a promising approach to identify biomarkers that are potentially translatable into clinical practice for diagnostic and prognostic purposes. Dysregulated transcripts identified at candidate gene and genome-wide levels provide us with new insights into biological mechanisms of depression. However, similar to gene expression studies in post-mortem brain, the findings in peripheral tissues lack consistency. In a recent systematic review, we re-analysed the results obtained from 10 transcriptomic studies in depression and showed that only 2.8% of genes (21 out of 752) identified as significantly differentially expressed in the periphery between depressed patients and healthy control subjects were replicated. Although a low level of replication at an individual gene level is discouraging, we made the observation that various factors of the immune response were consistently dysregulated. Among these 21 replicated genes, three were involved in the immune

system: *IFIT3*, which is involved in the type 1 interferon signalling pathway was found upregulated in MDD and downregulated in postpartum depression; *STAT3*, a transcription factor that mediates cellular responses to interleukins and *SEMA3C*, which is also known to be involved in the immune response, were found upregulated in MDD (Ciobanu et al., 2016). This provides further evidence for an involvement of peripheral inflammation in the pathophysiology of depression.

At this stage it is unknown how well the peripheral markers are predictors for brain inflammation (and vice versa). Despite a substantial overlap between brain and peripheral blood transcriptomes (Liew et al., 2006), the extent at which peripheral inflammation is a reflection of CNS inflammation in depressed individuals is not fully understood. One of the challenges in biomarker research for psychiatric disorders is to agree on the 'best' peripheral tissue source. While separated cell types or stimulated blood provide better resolution of signal compared to unstimulated blood, dysregulation of inflammatory and other immune-related genes is detectable even in unstimulated whole blood. For example, elements of disrupted immune signalling were found in unchallenged monocytes (Carvalho et al., 2014), PBMCs (Belzeaux et al., 2010, Savitz et al., 2013, Segman et al., 2010), dermal fibroblasts (Garbett et al., 2015b), whole blood after lipopolysaccharide (LPS) stimulation (Spijker et al., 2010), as well as in unstimulated whole blood (Jansen et al., 2016) from MDD patients. Although the aforementioned studies not only differ in cell type, but also in type of depression (postpartum, melancholic, induced by INF-α treatment), type of study design (targeted vs. genome-wide), and among cohort characteristics (age groups, medications), there was some agreement in findings on gene expression markers of inflammation between the studies .For instance, by studying a panel of 47 inflammatory-related genes Carvalho et al. (2014) found that one of two identified clusters, consisting primarily of pro-inflammatory mediators *(IL1A, IL1B, IL6, PTX3, PDE4B, PTGS2,* and *TNF),* were upregulated in monocytes of 47

patients with melancholic depression. This was somewhat consistent with genome-wide findings of a dysregulated functional network centred on differentially expressed *TNF* in PBMCs from 21 current and re-current moderately to severely affected MDD cases (Savitz et al., 2013). However, the latter study underscored that differentially expressed immune players were functionally linked with non-differentially expressed *NF-kß, TGFß,* and extracellular signal-regulated kinase *(ERK)*, indicating that differential expression analysis might be a suboptimal option for detecting complex gene-gene interactions in depression. In another study, one of the TNF receptors, *TNFRSF1* together with interferon-induced proteins *IFIT1, IFIT2*, and *IFIT3* and 8 other genes involved in immune response (*HELLS, HIST2H2B, GBPI, IGJ, SERPING, LOC44203, CXCL10, EREG)* were found differentially expressed in PBMCs of 9 patients with postpartum depression (Segman et al., 2010). In contrast, among 40 studied candidate genes including *TNF*, *IL1B, IL2, IL4, IL6, IL8*, and *IL10,* only the anti-inflammatory cytokine *IL10* was statistically significantly elevated in PBMCs of 11 individuals suffering from a severe melancholic depressive episode (Belzeaux et al., 2010). To overcome low signal to noise ratio of gene expression in basal blood, Spijker et al. (2010) used LPS challenged whole-blood cells from 21 individuals with a single MDD episode. They observed LPS-induced gene expression, among which there were several cytokines, such as *TNF, NF-kB, IL1, IL6,* and *IL10*. Although none of these genes displayed a differential expression level between MDD patients and control subjects, six out seven genes in a proposed diagnostic signature of depression are related to the immune system and deal with cellular proliferation (*CAPRIN1, PROK2, ZBTB16)* and differentiation *(CLEC4A, KRT23, PLSCR1).* Environmental influences, like lifestyle and medication use, can confound gene expression findings. Dermal fibroblasts were proposed as an alternative experimental model to study depression-specific gene expression alterationsas this cell type isless dependent of environmental influences (Garbett et al., 2015a). The authors argue that

after several rounds of fibroblasts' division in the cell culture, many epigenetic changes disappear over time, leaving a more "pure" genetic model at hand. Findings in fibroblasts were consistent with previously described observations in other cell types, pointing to disrupted molecular pathways related to cell-to-cell communication that are known to play a role in the adaptive and innate immune system. A set of 13 PR-qPCR validated immune-related genes was suggested to be associated with life style-independent and medication-free status in depression (*CD74, HLA-DRA, HLA-DQB1, IL11, HLA-DPA1, S100B, HBEGF, HLA-DPB1, HLA-DQA1, MET, PCDH10, TNF19, GSTT1).*

Table 5.2. Studies on the dysregulation of immune genes in peripheral tissues

Citation	Tissue/ Cell Type	Platform/Genes Studied	Diagnostic Criteria	Sample Characteristics	Medications	Main Findings
Carvalho et al. (2014)	Monocytes	47 Inflammatory-related genes RT-qPCR	DSM-IV-TR HAMD 17 SCID-1	47 Melancholic MDDcases (age range 32–82, 20% males) 42 Controls	Medication-free for at least 1 week	34 Genes upregulated, 2 downregulated Cluster analysis: First cluster—upregulated pro-inflammatory mediators (*IL1A, IL1B, IL6, PTX3, TNF, PDE4B*, and *PTGS2*)
Garbett et al. (2015)	Dermal fibroblasts	Affimetrix HT HG-U133^{+}	DSM-IV-TR	16 Current MDE cases Mean age 34.9, 4M:12F 16 Controls	Not relevant	13 "immune" genes were dysregulated: *CD74, HLA-DRA, HLA-DQB1, IL11, HLA-DPA1, S100B, HBEGF, HLA-DPB1, HLA-DQA1, MET, PCDH10, TNF19, GSTT1*
Felger et al. (2012)	PBMCs	Illumina HT-12	DSM-IV SCID MMSE MADRS	21 HCV patients 12M:9F	No antidepressants for at least 4 months prior 12 weeks of INF-α/ribavirin treatment	IFN-α treated patients with high depression scores showed upregulation in INF-α and *AP1* signaling and reduced prevalence of *CREB/ATF* motifs
Hepgul et al. (2016)	Whole blood	Affimetrix HG 1.1 ST	DSM-IV IDS	58 HCV patients 20 Patients developed IFN-α induced major depressive episode	24 weeks of INF-α treatment	More genes (506) were modulated in patients who developed depression with enrichment in inflammation- and oxidative stress-related pathways
Jansen et al. (2016)	Whole blood	Affimetrix U219	DSM-IV	882 Current MDD cases 331 Controls	Antidepressants used as covariates	Upregulation of IL-6 signaling and downregulation of NK-cell cytotoxicity pathways
Mostafavi et al. (2014)	Whole blood	Illumina HiSeq 2000	DSM-IV	463 Recurrent MDD cases 459 Controls	Antidepressants used as covariates	No significant single-gene association after multiple-testing correction; however, there was increased expression of interferon α/β pathway

Continued

Citation	Tissue/ Cell Type	Platform/Genes Studied	Diagnostic Criteria	Sample Characteristics	Medications	Main Findings
Savitz et al. (2013)	PBMCs	Illumina Human HT-12 v4 fMRI scanning	DSM-IV-TR HDRS	21 MDD (current or recurrent moderate to severe) cases And eight adults with BP in a current MDE; age 35 ±10, 32% males 24 Controls	No psychotropic medications for at least 3 weeks (eight for fluoxetine)	12 Protein-coding *genes* (*ADM, APBB3, CD160, CFD, CITED2, CTSZ, IER5, NFKBIZ, NR4A2, NUCKS1, SERTAD1*, and *TFN*) were dysregulated. One functional network is centered around *TNF*; inflammatory genes correlate with gray matter volume of the hippocampus and caudate and thickness of subgenual ACC
Segman et al. (2010)	PBMCs	Affimetrix Human Gene Chip Exon ST 1.0	Edinburgh Postnatal Depression Scale	9 PD 10 Controls	Antidepressant naive	73 Differentially expressed genes, 71 downregulated, 12 of them involved in immune response (*HELLS, HIST2H2B, GBPI, IFIT3, IGJ, SERPING, IFIT1, IFIT2, LOC44203, CXCL10, TNFRSF1*, and *EREG*)
Spijker et al. (2010)	Whole blood ex vivo LPS-stimulated	Agilent	DSM-IV IDS-SR30	21 Single or recurrent MDD episode cases Mean age 42.6 ±11.5, M:F 21 Controls Validation: 13 MDD cases 14 Controls	No current antidepressants or benzodiazepines	Seven genes were proposed as diagnostic signature of depression, six of which are related to immune system and deal with cellular proliferation (*CAPRIN1, PROK2*, and *ZBTB16*) and differentiation (*CLEC4A, KRT23*, and *PLSCR1*)

LPS, lipopolysaccharide; *CRS*, chronic restraint stress; *HCV*, chronic hepatitis C virus; *HDRS*, Hamilton Depression Rating Scale; *MDE*, Major Depressive Episode, postpartum depression; *IDS-SR30*, The Inventory of Depressive Symptomology; *MINI*, Mini-International Neuropsychiatric Interview; *PHQ-9*, Patient Health Questionnaire-9; *MMSE*, mini-mental state examination; *MADRS*, Montgomery-Asberg Depression Rating Scale; *IDS*, Inventory of Depression Symptomatology.

While we report on some agreement of gene expression markers of inflammation in the periphery across studies, the majority of immune-related genes have not been replicated. One of the major factors that may have led to the disagreement between the studies is low statistical power. Small sample sizes, ranging from only 9 to 47 depression cases (Table 5.2), is one of the main limiting factors affecting statistical power to detect and replicate dysregulated transcripts. To overcome this limitation, (Jansen et al., 2016)

performed the largest to date study on 882 patients with current MDD. Using both differential expression and co-expression clustering (WGCNA) (Zhang and Horvath, 2005) methods and accounting for 16 demographic and technical covariates, they identified that MDD is characterised by upregulated IL-6 signalling (*IL6R, STAT3, MAPK14, RXRA*) and downregulated NK cell activation (*GZMB, KLRK1, PRF1, SH2D1B, KLDR1, NFATC2, IL2RB, CALM1, NCALD*). Longitudinal analysis at 2-year follow-up showed that the levels of expression for 15% of genes out of the 129 genes identified in cross-sectional analyses were reversed in those who remitted after a previous depression episode. This indicates that transient gene expression patterns are detectable in peripheral blood and the results provide support for the potential success in the development of whole-blood gene expression-based biomarkers of depression. These results were also meta-analysed with a recent RNA-seq study, in which alone no significant association with depression on 463 self-reported MDD cases accounting for 39 covariates were found. Binding the two largest datasets derived from microarrays and RNA-sequencing technologies together, resulted in 12 differentially expressed genes at FDR<0.1 between MDD cases and healthy controls, 7 of which are known to be involved in activated immune signalling and oxidative stress (*CALM1, FCRL6, APOBEC3G, RAP2B, PIPOX, PRR5L*, and *KLRD1*), providing further support for the often reported peripheral inflammation in depression.

Taken together, an intensive search for peripheral gene expression biomarkers of inflammation in depression at both the candidate gene and transcriptome levels identified some promising candidates, including *IL1B, TNF, IL6*, *INF, CREB1* and *NF-kB*, the field is far away from claiming an immune signature of depression. Further understanding of gene expression alterations in depression with a particular focus on immune-related pathways will help us to better understand the role of immune system in depression.

In our previous study, employing co-expression network approach, we found that 13 biological immune-related pathways are significantly altered (activated or inhibited) in older adults suffering with recurrent MDD. However, the exact mechanisms of this dysregulation requires further investigation. Targeting these immune related pathways at individual gene level to determine what genes are differentially expressed in depression is complementary to candidate gene and transcriptome-wide approaches. In this study, we select a pathway-targeted subset of biologically relevant genes that can provide us with more granular understanding on immune dysregulation in depression. To select a subset of targeted immune genes for further analyses, we (1) extracted the full list of genes that are known to be involved in a given pathway, (2) removed redundant information about genes involved in multiple pathways, and (3) conducted differential gene expression analysis between individuals with recurrent MDD and those without using the OATS cohort of elderly community-dwelling individuals.

METHODS

Candidate gene selection and mapping gene names to Illumina probes

The 13 KEGG pathways associated with recurrent MDD in the elderly are listed in the Table 5.3. In total, there were 1,869 genes involved in these pathways with 30.3% (576) of them involved in multiple pathways. After removing duplicated values 1,302 individual genes across 13 KEGG pathways remained for further analysis. For viewing KEGG pathway diagrams with “depression” genes marked up, refer to Supplementary Materials for (Ciobanu et al., 2018b). The 1,302 genes of interest were mapped to the 2,085 corresponding Illumina HumanHT-12 v4 probes.

Table 5.3. Thirteen KEGG pathways associated with recurrent MDD (Ciobanu et al., 2018b)

Pathway name	KEGG ID	Total N	Status in MDD	FDR
Fc gamma R-mediated phagocytosis	hsa:04666	91	Activated	0.001
Shigellosis	hsa:05131	65	Activated	0.006
mTOR signalling pathway	hsa:04150	153	Activated	0.006
Natural killer cell mediated cytotoxicity	hsa:04650	131	Inhibited	0.008
Antigen processing and presentation	hsa:04612	77	Activated	0.01
Herpes simplex infection	hsa:05168	492	Activated	0.01
Insulin signalling pathway	hsa:04910	137	Activated	0.03
Pathogenic Escherichia coli infection	hsa:05130	55	Activated	0.03
RNA degradation	hsa:03018	79	Activated	0.04
Epstein-Barr virus infection	hsa:05169	201	Activated	0.04
Protein processing in endoplasmic reticulum	hsa:04141	165	Activated	0.04
Apoptosis	hsa:04210	136	Activated	0.04
T cell receptor signalling pathway	hsa:104660	101	Inhibited	0.046

Pathway name, KEGG ID, and Total N (total number of genes known to be involved in a given pathway) refer to the KEGG database; Status in MDD (activated or inhibited), MDD N (number of genes found differentially expressed in MDD), and FDR (at 5%) refers to the previous findings (Ciobanu et al., 2018b).

Cohort characteristics

To conduct the differential expression analyses on immune related genes in recurrent MDD we utilised the Older Australian Twins Study (OATS). The primary aim of OATS is to investigate healthy brain ageing in older twins (65+ years) (Sachdev et al., 2009). Depression data were collected at three time points with 2-year intervals between assessments. Blood samples for gene expression analyses were collected at Wave 3,

four years after baseline. The total sample size available for further analyses N=186 (recurrent MDD = 7). Informed consent was obtained for all participants and study procedures were explained prior to study commencement. Further details on cohort characteristics can be found in (Ciobanu et al., 2018b).

MDD definition

MDD was assessed by two well-validated self-report, and two clinical interview-based, measures of depression including the Geriatric Depression Scale (GDS-15) (cut-off ≥ 6) (Yesavage et al., 1982), the Patient Health Questionnaire (PHQ-9) (cut-off ≥ 10) (Kroenke K and R., 2002), the Neuropsychiatric Inventory (NPI) (depression sub-scale) (Cummings et al., 1994), and the Mini International Neuropsychiatric Interview (M.I.N.I.) (Sheehan et al., 1998). The recurrent MDD cases were defined as participants who were identified as clinically depressed by at least two of the above depression measures, according to DSM-IV criteria, at more than one study Wave.

Gene expression data pre-processing

Gene expression data for 186 whole blood samples collected in PAXgene tubes were generated using the Illumina Whole-Genome Gene Expression Direct Hybridization Assay System HumanHT-12 v4 (Illumina Inc., San Diego, CA, USA) according to standard manufacturer protocols. After rigorous QC (RIN<6), pre-processing, filtering, and adjustments for known (RINs, age, sex) and latent non-biological variables, 11,685 probes for 186 individuals aged 69 and over from the OATS cohort (age range 69.4 – 93.5, mean age 75.9±5.3, 72 males and 114 females) were available for downstream analyses.

Differential gene expression analysis

To assess differential gene expression, DGE, between individuals diagnosed with

recurrent depression and those without this diagnosis, we fit gene-wise linear modelling implemented in *limma R* package. The linear modelling was performed in a row-wise fashion, with regression coefficients and standard errors estimating the comparisons of interest. The fitted model and contrast matrix were used to compute $\log_2$-fold-changes and *t*-statistics for the contrasts of interest. For each coefficient in the contrast, empirical Bayes moderated *t*-statistics and their associated *P*-values were computed to assess the significance of the observed expression changes.

RESULTS

For the 1,302 individual genes that were identified using KEGG database there were 963 Illumina corresponding probes (750 unique genes) in our OATS gene expression dataset available for DGE analysis. The 102 probes (90 unique genes) differentially expressed at nominal $p < 0.05$ (Appendix C). After FDR correction for multiple testing, two probes corresponding to two splice variants of the *CD14* gene (ILMN1740014 and ILMN239644) remained significant at FDR 5%. Both probes were downregulated with comparable effect (logFC = - 0.7 and - 0.64 accordingly). Five more genes, namely *LSM2, ACTB, ATP6V1B2, CFP* and *PAK 1*, can be considered significantly dysregulated at relaxed FDR 25% threshold (Figure 5.1, Table 5.4, Table S5.1).

Figure 5.1. Differentially expressed genes in recurrent MDD. The volcano plot displaying log fold changes on the x-axis versus a measure of statistical significance on the y-axis (here the significance measure is the posterior log-odds of differential expression (B-statistics). The top six genes are named, statistically significant gene at FDR 5% marked with asterisk.

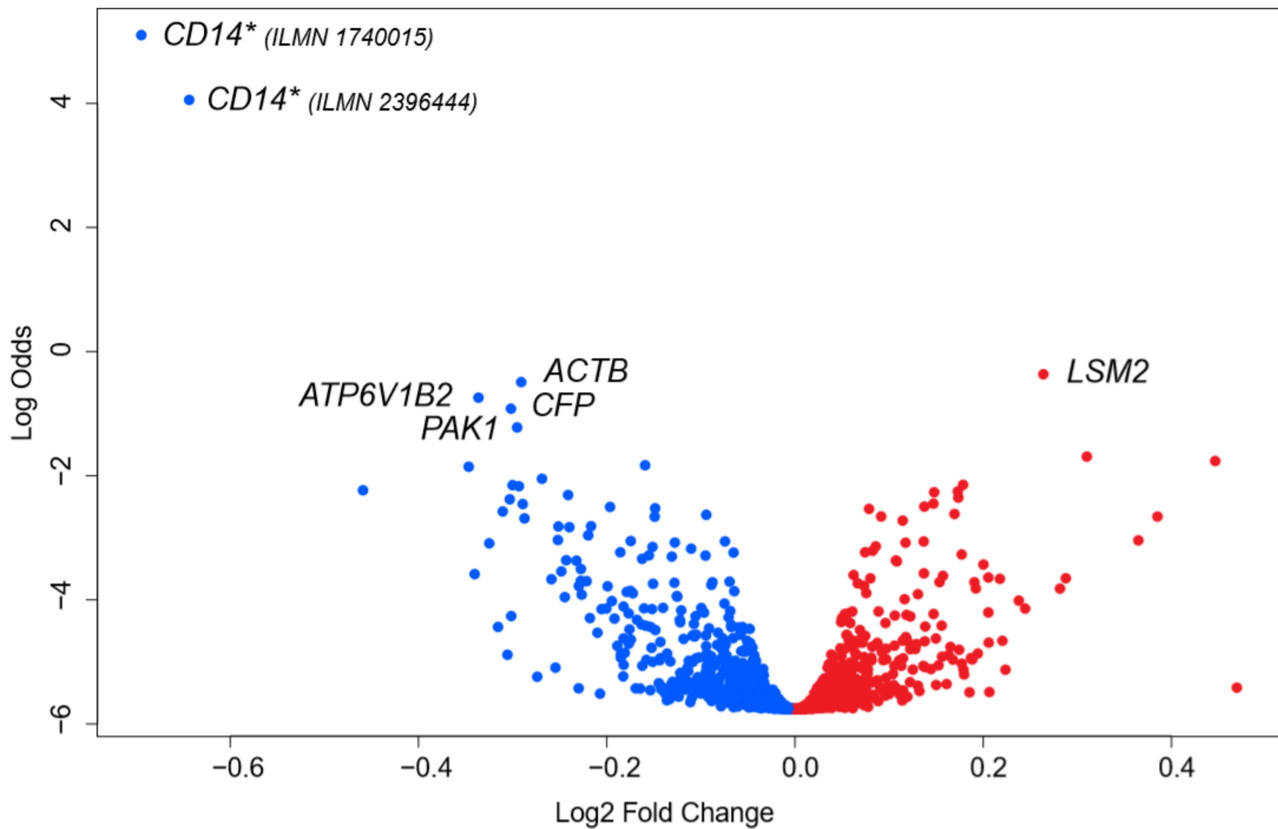


Table 5.4. The top 6 differentially expressed genes in recurrent MDD

Official symbol	Illumina probe	ENTREZ	Official gene name	logFC	p-value	FDR
CD14	ILMN_1740015	929	CD14 molecule	-0.7	9.51E-07	0.00092
CD14	ILMN_2396444	929	CD14 molecule	-0.64	3.19E-06	0.002
LSM2	ILMN_2070300	57819	LSM2 homolog, U6 small nuclear RNA and mRNA degradation associated	0.26	0.0006	0.16
ACTB	ILMN_2152131	60	actin beta	-0.29	0.0007	0.16
ATP6V1B2	ILMN_1787705	526	ATPase H+ transporting V1 subunit B2	-0.34	0.0009	0.17
CFP	ILMN_1658121	5199	complement factor properdin	-0.3	0.001	0.18
PAK1	ILMN_1767365	5058	p21 (RAC1) activated kinase 1	-0.3	0.002	0.22

DISCUSSION

In our previous study (Chapter 3) using co-expression network analyses we reduced dimensionality of transcriptome data in a functionally relevant manner and identified immune- and pathogen-related pathways associated with recurrent depression. To gain further insight into immune dysregulation in recurrent MDD at gene expression level we employed a targeted candidate pathway approach. After selection of 13 immune-related pathways associated with recurrent depression in the elderly, we found two mRNA variants that encode for the *CD14* molecule to be significantly downregulated in individuals with recurrent MDD compared to those without (FDR < 0.05) in independent cohort. This finding confirms previous observations that innate immune response-

activating signal transduction mechanisms are altered in depression and suggests that *CD14* may represent a novel putative biomarker for recurrent MDD.

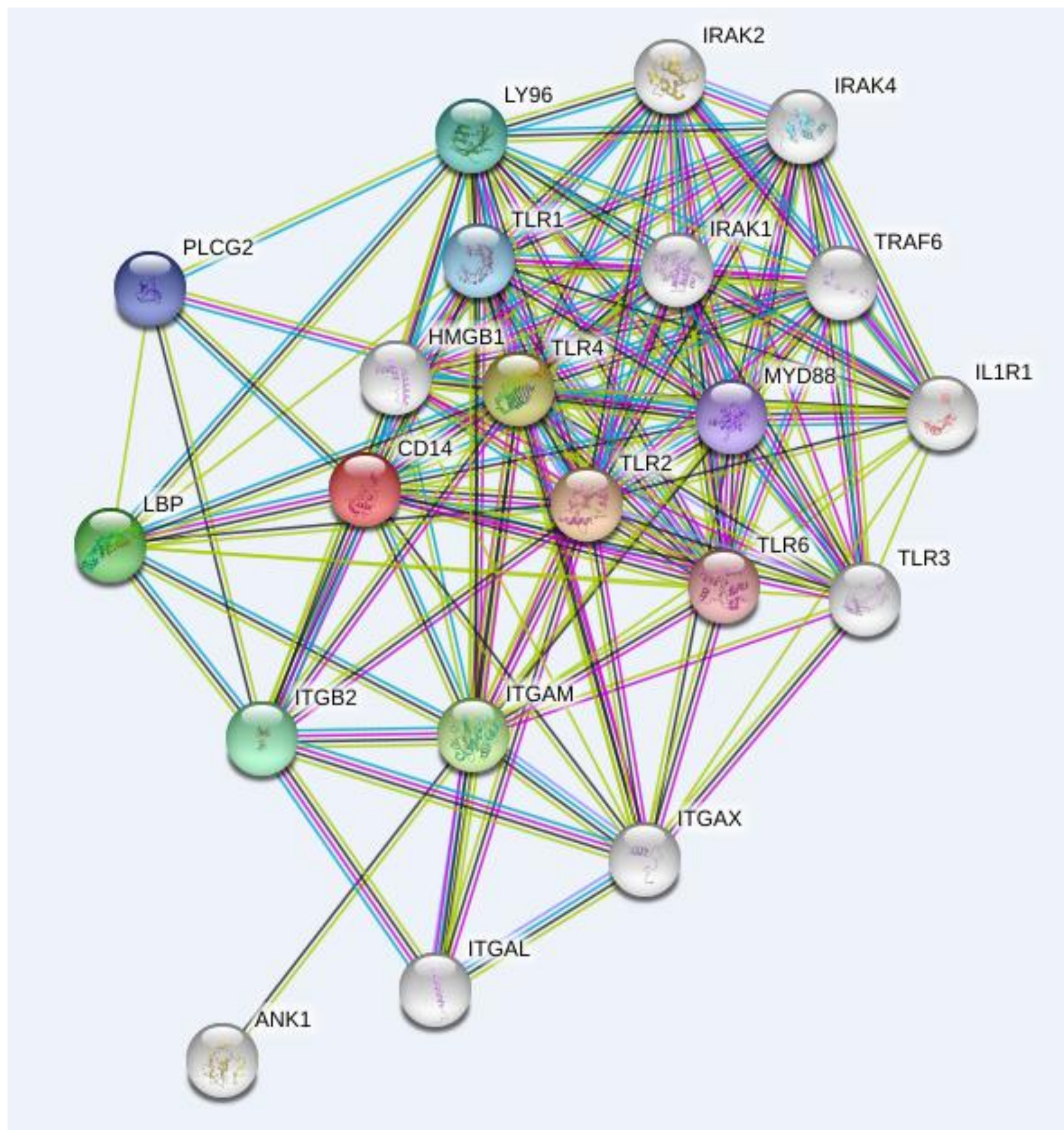
The role of CD14 in innate immune responses and oxidative stress

Cluster of differentiation antigen 14 (CD14) is a glycosylphosphatidylinositol (GPI)-anchored receptor known to serve as a co-receptor for several Toll-like Receptors (TLRs) both at the cell surface and in the endosomal compartment. *CD14* is mainly expressed in peripheral tissues, including whole blood; it also expressed in the brain (hippocampus, cortex, nucleus accumbens etc.) although in a much smaller quantity (https://gtexportal.org/home/gene/CD14). CD14 is known as a critical upstream regulator of the host's inflammatory response to pathogenic challenge; it influences the intensity and duration of inflammation to finely modulate NF-κB activity. While CD14 is a multifunctional molecule that participates in many immune-related biological pathways (Figure 5.2), current thinking emphasizes the role of CD14 in facilitating recognition of pathogens to initiate and orchestrate innate immune-mediated signalling events (Sahay et al., 2009, Zanoni and Granucci, 2013).

The role of *CD14* in oxidative stress is emerging. Recent studies show that CD14 has an impact on Reactive Oxygen Species (ROS) production when primed by *Escherichia coli* lipopolysaccharides in human leukocytes in vitro (Kabanov, 2019); activation of CD14-dependent innate immune response was also found causal to neuronal oxidative damage and dendritic degeneration in vivo (Milatovic, 2004). Therefore, if the expression of *CD14* is compromised, activation of an intracellular signalling pathway *NF-kB* and inflammatory cytokine production that are responsible for facilitating the innate immune system would be impaired leading to damage by ROS (Liu et al., 2012, Lau et al., 2015). Together with our finding on a role of the peroxisome lipid and ROS metabolism pathway in depressive disorder (Chapter 2), the results on altered *CD14* expression in recurrent MDD suggest

that dysregulated CD14-dependent innate immune cascade leading to accumulation of ROS-induced damage plays a role in recurrent MDD and could represent a recurrent MDD-specific biomarker and potential therapeutic target.

Figure 5.2. Interaction network for *CD14*. Network nodes represent proteins. Enrichment for top 10 GO terms is presented



GO ID	Description	FDR
GO:0002224	toll-like receptor signaling pathway	2.67E-30
GO:0002758	innate immune response-activating signal transduction	6.90E-29
GO:0002755	MyD88-dependent toll-like receptor signaling pathway	7.93E-23
GO:0050778	positive regulation of immune response	7.93E-23
GO:0050776	regulation of immune response	5.30E-22
GO:0031347	regulation of defense response	7.16E-22
GO:0006954	inflammatory response	3.49E-18
GO:0001819	positive regulation of cytokine production	1.52E-17
GO:0006955	immune response	1.90E-17
GO:0042116	macrophage activation	4.11E-17

The network was generated using STRING software https://string-db.org

The role of CD14 in age-related alterations in innate immune response and ROS production

Aging is associated with dysregulation of multiple components of the immune system that results in a decrease of adequate response to pathogens and increase in susceptibility to infections (Metcalf et al., 2015, Weiskopf et al., 2009). Furthermore, oxidative stress is a hallmark of ageing (Liguori et al., 2018). Considering that *CD14* is an important modulator of both pathogen-activated innate immune response and production of ROS, our findings on dysregulated *CD14* within the elderly can be partially attributed to the aging process itself. However, it can also indicate that within the elderly population depression could be stratified according to particular innate immune and oxidative stress mechanisms coordinated by *CD14*.

While the population design employed in this study (individuals with recurrent MDD vs those without) may suggest that the results obtained have a stronger potential to be translated into actionable clinical insights, the unbalanced ratio of those affected (minority) with those not (majority) used in differential expression analysis can potentially produce unreliable results. Therefore, replication of these findings on a larger cohort employing case-control study design is required.

CONCLUSIONS

Considering a central coordinating role of *CD14* in orchestrating the host's inflammatory response to pathogenic challenge and association of downregulated *CD14* with recurrent MDD in the elderly, our results suggest that CD14-mediated immune response is impaired in recurrent MDD. However, replication of these results on a larger cohort is warranted in

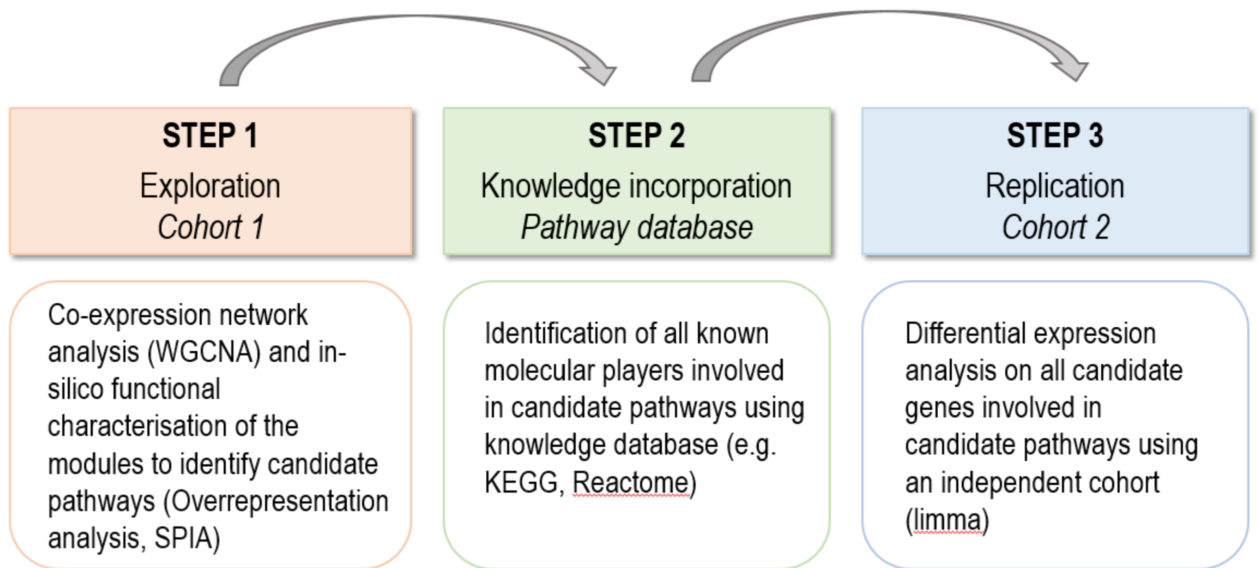
addition to cellular based experiments.

A NOVEL CANDIDATE PATHWAY FRAMEWORK

Navigating our way towards a greater understanding of the pathophysiological mechanisms in depression requires a combined focus on the forest (global focus, transcriptome-wide view) and focus on the trees (local focus, individual genes view). It's becoming clear that whilst it is important to identify individual players via for example differential gene expression analyses, that can be utilised as biomarkers and therapeutic targets, this cannot be done without taking a global perspective on a whole system of complex interactions between the genes. Co-expression network (WGCNA) and differential gene expression represent two different but complimentary analytical strategies. While co-expression network takes a global systems approach to identifying disrupted biological pathways in a disease, differential gene expression is a powerful method to pinpoint individual players. Combining these two methodologies in a workflow provides a much more comprehensive analytical approach to gain novel insights into the underlying biological mechanisms of MDD.

We propose a novel candidate pathway framework that incorporates both global and local perspectives. The framework consists of the three major steps: (1) transcriptome co-expression network construction (WGCNA) to identify biological pathways associated with depression (an example can be found in Chapter 3), (2) identification of all genes known to be involved in given candidate pathways using pathway database (KEGG), (3) differential gene expression analysis of genes involved in candidate pathways on an independent cohort (limma, R) (an example can be found in Chapter 5). Figure 5.3 outlines the major steps of the candidate pathway framework.

Figure 5.3. A candidate pathway framework



This framework allows for a comprehensive investigation of transcriptome signature in depression from a complex network of interactions at genome-wide level to individual genes and includes replication process on an independent cohort. Using proposed candidate pathway framework, we were able to identify pathogen- and immune-related molecular pathways implicated in recurrent MDD (Chapter 3) and identified *CD14* as one of the central players that appear to be responsible for orchestration of these pathways (Chapter 5). This suggests that a candidate pathway framework is useful for detecting relevant signals in highly multidimensional transcriptome data with poor signal-to-noise ratio and can be further utilised for studying gene expression signatures of depression or other heterogeneous conditions.

CHAPTER 6

Summarising the findings and outlining future directions

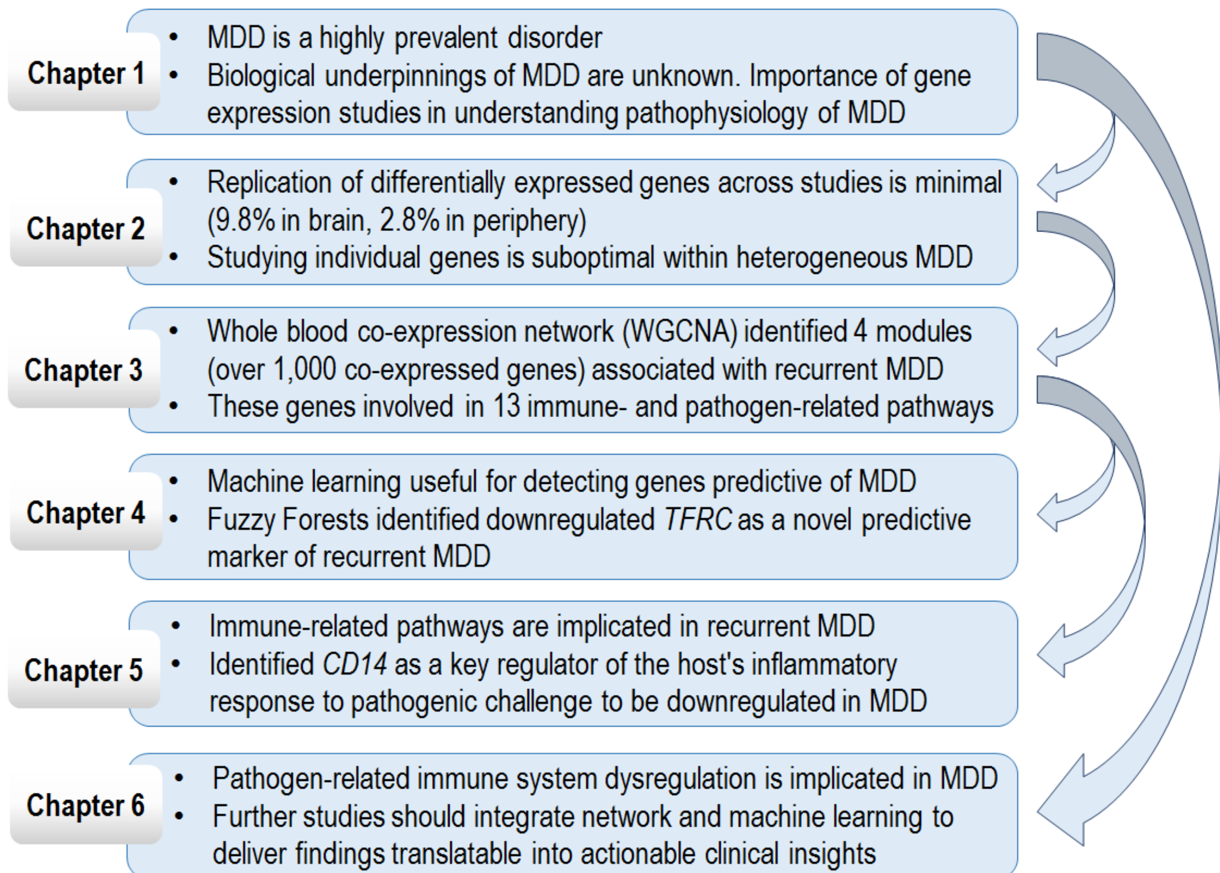
Elucidating transcriptome signature of depression is critical for understanding biological underpinnings of depression. In general, transcriptome research in MDD Over the last 10 years of has been challenging due to multiple limiting factors of technical and biological nature, which we extensively discuss throughout this manuscript. In particular, for this study some technical aspects and experimental limitations will be discussed.

There are a number of highly significant confounders, which would potentially influence the replicability of expression levels of specific genes, and therefore, the interpretability of the results obtained in this work. Of pivotal relevance is the pharmacological treatment of patients included in each study, and the potential impact of confounding of pharmacological treatment received by MDD patients. It is quite possible that differences in gene expression between MDD patients and healthy controls reported in previously published studies (examined in Chapter 2) are influenced by those which are regulated, consequently, of the treatment received, and that differences in DEGs observed between studies may reflect different treatments received by patients in each study. Differences in treatments may also explain the lack of replicability across the studies observed (Ciobanu 2016). Our main experimental cohorts, namely sMAS and OATS, utilised in Chapters 3, 4 and 5 for transcriptome data analyses in MDD also suffered from a lack of information regarding pharmachological treatments taken by patients prior to blood collection, which might have affected the obtained results. Future transcriptome studies should consider pharmacological treatment at the stage of study design to facilitate adjustment for this important confounder at data analysis step.

In addition, the different sensitivities of the methods used to capture gene expression (i.e. microarray vs RNA-seq) can have an effect on the results obtained. For example, when using microarray data, low abundance genes are usually filtered out before the analysis due to lack of sensitivity to reliably detect them, while RNA-seq has higher sensitivity for low abundance genes and these genes can be included in statistical analysis. Furthermore, microarrays cannot capture alternative mechanisms of functional gene regulation (e.g. alternative splicing isoforms or non-coding transcripts not typically captured by microarray), while RNA-seq technology can, which may also be relevant in the molecular mechanisms of depression.

To overcome some of the limitations observed in the previous studies (Chapter 2) that may have affected the progress in the field, we applied several novel statistical and machine learning techniques, such as co-expression network (Chapter 3), fuzzy forests (Chapter 4), and candidate pathway (Chapter 5) on two large population cohorts of older adults. We found that (1) immune dysregulation characterised by thirteen pathogen- and immune-related molecular pathways appear to have an important role in recurrent MDD (Ciobanu et al., 2018b), the downregulated *TFRC* can predict recurrent MDD in the elderly, and therefore can potentially be used as a predictive marker of the disease for the specific subgroup of patients (Chapter 4, paper under review), and (3) *CD14*-coordinated innate immune signalling is associated with recurrent MDD and can represent a potential novel therapeutic target (Chapter 5). Future directions are outlined for the importance of emerging methodologies and suggest several approaches that have a potential to facilitate biological findings into actionable clinical insights. Figure 6.1 illustrates summarised findings of the PhD thesis.

Figure 6.1. Summary of the findings on investigation of transcriptome signature of depression



In summary, while supporting previous findings on the immune system being involved in depression, our work provides converging evidence that pathogen-related immune dysregulation plays a role in recurrent MDD in older adults and highlights two candidate genes (*TFRC* and *CD14*) as potential peripheral markers of the disorder.

The next section outlines how advanced analytical strategies can enhance biological psychiatric research for stratification of subgroups via molecular signatures and treatment response biotypes. This will further advance our understanding of disorder mechanisms and accelerate transition into an era of personalized treatments in psychiatry.

GENE CO-EXPRESSION NETWORK AND MACHINE LEARNING APPROACHES TO TRANSLATE BIOLOGICAL FINDINGS INTO ACTIONABLE CLINICAL INSIGHTS

Despite the success of collaborative international efforts to identify genetic variants involved in depression, the biological underpinnings of complex psychiatric traits, including depression, remain elusive. Increasing evidence suggests that depression is the result of complex interactions between genomic variations, epigenetic modifications, and other regulatory mechanisms involved in gene expression. Therefore, the transcriptome, representing a nexus of genetic and environmental interactions, can be seen as an essential biological layer of information for studying molecular dysregulations in mental disorders. Transcriptomics can be used for diagnostic purposes to differentiate disease from healthy controls, differentiate disease stages and identify subgroups of patients exhibiting different biological signatures within diagnosis. It also allows measuring the influence of drugs on the transcriptome, which can be useful in getting insights on molecular mechanisms of drug's action and in predicting treatment response. Using co-expression network analysis-based methods, disease-relevant clusters of co-regulated genes can be identified and further integrated with genetic and epigenetic data for a comprehensive investigation of biological underpinnings of depression. In this chapter, we provide an essential guide to the co-expression network approach as an important statistical tool that can enrich the understanding of disrupted molecular processes in psychiatric disorders. Furthermore, given the complex system structure inherent in psychiatric disorders, including depression, statistical learning frameworks that can translate these findings into actionable clinical insights are required. A number of emerging methodologies that address this problem are explored, including weighted gene co-expression network analysis, differential co-expression analysis, biclustering, and regularised machine learning.

Emerging evidence suggests that interactions between both genes and the environment can have a large impact on the phenotypic variability of psychiatric disorders (Kubota et al., 2012). Given that gene expression is a product of genetic effects, environmental influences and epigenetic modifications, studying alterations of mRNA levels in disease represents a promising approach to further our molecular understanding of depression. However, a major challenge has remained. How is it that we go about measuring and identifying the expression of which genes are altered in disease specific cells, tissues and brain regions?

Early studies measuring the differential expression of candidate genes pointed towards many potential targets for depression. However, the replication of these findings has been limited due to methodological differences, inconsistencies in the diagnostic measures used, and underpowered study designs (Drago et al., 2007). Nevertheless, advances in high-throughput technologies have helped facilitate a shift from hypothesis-driven approaches, to less biased data-driven methodologies. A substantial body of research has applied microarray and RNA-seq technologies to investigate disease-associated genome-wide expression alterations in both brain and peripheral tissues. This work has been demonstrated in schizophrenia (Sanders et al., 2017, Cattane et al., 2015), bipolar disorder (for review, Seifuddin et al. (2013)), and MDD (for review, (Ciobanu et al., 2016)). While these findings have improved our understanding of the pathophysiological mechanisms in psychiatric disorders, replication of identified candidates has remained a challenge. One proposed reason for the lack of consistent results between studies is attributed to the molecular complexity of psychiatric disorders, which commonly, is further compounded by the small sample sizes often employed in this research. The capacity to detect multiple small effects from single molecules at the transcriptome-wide level places large sample size requirements on a study. Given the complex genetic architecture of psychiatric illness and the functional interdependencies between genes, emergent clinical

phenotypes are likely to reflect the interactions within a complex network of molecular processes. Therefore, capturing the most important genes that orchestrate the molecular cascades leading to psychiatric pathophysiology is of utmost importance. However, given the methodological constraints discussed above in traditional analysis, alternative methodologies are needed to further progress the field. A range of these methodologies will be discussed below.

DIAGNOSIS OF DEPRESSION

Co-expression network analysis to differentiate disease from healthy controls

Gene co-expression network analysis is a technique used to quantify the linear co-expression of multiple genes in relation to disease diagnosis. This methodology allows for effective dimensionality reduction of transcriptome data (decreasing the number of independent variables in a dataset), as well as, clustering of interacting genes associated with a phenotype. It can be argued that if two (or more) genes are co-expressed, then the mechanisms regulating their expression must be either the same, or at least similar. Therefore, it is inferred that co-expressed genes are functionally related, and if associated with a disease status, are important contributors to a clinical phenotype.

Weighted gene co-expression network analysis

Gene clustering is a technique used in co-expression analysis that allows for the identification of sub-networks of convergent biological pathways. The Weighted Gene Co-expression Network Analysis (WGCNA) is one of the most established techniques used in gene clustering. This technique utilises both microarray and RNA-seq data (Zhang and Horvath, 2005). WGCNA organizes transcriptome data by defining genes as nodes and the relationships between nodes as edges. Biological networks tend to have a

hierarchical structure, such that their nodes can be clustered together into fewer modules of highly interconnected genes. Inter-modular connectivity reflects a higher-order structure of biological relationships within a gene network, while intra-modular connectivity can identify which genes behave as central hubs in the modules. In co-expression networks, hubs are highly connected genes; therefore, being a hub is an indication of the importance of a gene in a module. Hubs are likely to be key molecular drivers that determine co-expression. Evidence also suggests that they may help to interpret a module as they likely reflect its predominant biological role.

WGCNA has been successfully applied in many areas of medical research, including psychiatry. For example, WGCNA has been used to differentiate disease cases from healthy controls and identify brain based immune system dysregulation in schizophrenia (Mistry et al., 2013), postsynaptic density implicated in the pathogenesis of bipolar disorder (Akula et al., 2015), peripheral inflammation in depression (Malki et al., 2013) and transcriptional and splicing dysregulation as underlying mechanisms of neuronal dysfunction in autism spectrum disorder (ASD) (Voineagu et al., 2011). Co-expression network analysis has also been applied to explore overlap and specificity across different disorders compared to healthy controls. For example, comparing gene co-expression patterns between adult and childhood attention-deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), major depressive disorder (MDD) and healthy controls, it has been found that immune system dysregulation is involved in both MDD and adult ADHD and is inversely correlated with a disease status (de Jong et al., 2016). Using WGCNA, neuron differentiation and development pathways in cerebral cortex have been discovered as potential contributors to the etiologies for both schizophrenia and bipolar disorder (Chen et al., 2013); while distinct molecular signatures have been found for ASD and intellectual disability (ID) (Parikshak et al., 2013). Recent work from Gandal et al. (2016) utilised WGCNA to compare molecular phenotypes across 5 major psychiatric

disorders, including autism (ASD), schizophrenia (SCZ), bipolar disorder (BD), depression (MDD), and alcoholism (AAD). From this comparison, authors were able to identify a clear pattern of shared and distinct gene-expression perturbations across all conditions. It has been found that neuronal gene co-expression modules were downregulated across ASD, SCZ, and BD, and astrocyte related modules were predominantly upregulated in ASD and SCZ.

Differential co-expression network analysis to study differences among gene interconnections

Differential co-expression analysis is a tool that is used to investigate the differences among gene interconnections. This is achieved by calculating the expression correlation change of each gene pair between conditions. Genes that are differentially co-expressed between different conditions are more likely to be regulators, thus, are likely to explain differences between phenotypes (Amar et al., 2013). By comparing the regulatory differences between cases and controls, specific differential networks of genes can be identified in psychiatric disorders. This methodology can be utilized to better understand the dynamic changes in gene regulatory networks in psychiatric illness, as well as comparing network properties across disorders. Thus, Xu et al. (2015) showed that mechanisms underlying MDD and subsyndromal symptomatic depression (SSD) were actually different. Authors found that there was no overlap between the MDD and SSD differentially regulated genes. Furthermore, the authors also found that Venlafaxine appeared to have a significant effect on the gene expression profile of MDD patients but no significant effect on the gene expression profiles of SSD patients. For more information on differential co-expression network analysis, see Hsu et al. (2015).

Biclustering for identifying subgroups of patients

Although gene co-expression clustering algorithms have proven useful for studying the molecular complexity of psychiatric disorders, the dominant clustering methodology discussed above is limited in its ability to detect gene expression patterns that are specific to subgroups of patients. An alternative approach called biclustering may be able to overcome this limitation. Biclustering is an alternative method for detecting differentially co-expressed genes between subgroups of patients, allowing for patient stratification into unique biological sub-groups. Biclustering algorithms perform clustering without the need for prior sample group classification, a beneficial characteristic given the uncertainty of diagnostic boundaries across psychiatric disorders. Using this approach, Cha et al. (2015) explored shared molecular basis of five neurodegenerative diseases (Alzheimer's disease, Parkinson's disease, Huntington's disease, multiple sclerosis, amyotrophic lateral sclerosis) and three psychiatric disorders (SCZ, BP, and ASD). The authors found that 4,307 genes were co-expressed in multiple brain diseases, whilst 3,409 gene sets were exclusively specified in individual brain diseases. Using the same approach in the field of oncology, Fiannaca et al. (2015) were able to identify unique molecular subgroups of breast cancer tumours based off of patients' miRNA expression profiles. Interestingly enough, all patients presented with the same clinical profiles. These findings are beneficial for clinical practice as they may allow for the molecular stratification of both diagnosis and treatment of disease, thus, allowing clinicians to tailor treatment strategies to individual patients. At present, biclustering methods show promise to help further elucidate the complexity and biological heterogeneity of complex psychiatric traits. For an extensive overview of the different biclustering algorithms available, refer to (Pontes et al., 2015).

PERSONALISED TREATMENTS

Gene co-expression networks for understanding treatment response biotypes

Understanding the biological basis of why some patients respond to certain therapies and others do not is essential for advancing personalised care in psychiatry. Due to the ability of gene co-expression analysis to approximate the complex interactions of biological information, it can be argued that this technique may help to inform the isolation of unique treatment response biotypes. Systematic characterisation of changes in gene co-expression patterns in responders vs. non-responders may provide biological grounds for developing predictive models that help to minimise therapeutic uncertainty. Furthermore, it may help to reveal novel mechanisms of action that remain unidentified in commonly used psychiatric drugs. Support for this notion has already been demonstrated in the field of oncology to delineate responders and non-responders to cancer treatment (Hsiao et al., 2016, Yang et al., 2014). In the field of psychiatry, co-expression network analysis of peripheral blood has identified immune-related pathways as important predispositions to antidepressant treatment response in MDD (Belzeaux et al., 2016). Blood-derived gene expression signature was found predictive of clozapine monotherapy in psychosis (Harrison et al., 2016). Furthermore, several gene co-expression modules in patient-derived lymphoblastoid cell lines were discovered as lithium-responsive, indicating widespread effects of lithium on diverse cellular signalling systems including apoptosis and defence response pathways, protein processing and response to endoplasmic reticulum stress in bipolar disorder (Breen et al., 2016). As these studies suggest, applying co-expression based network methods across different disorders and medications can help to yield important insights on the molecular interactions regulating treatment response in complex psychiatric traits.

Gene co-expression network for integrative analyses in psychiatry

Due to the multiple testing burden, large samples sizes are required to detect disrupted genes and pathways when multiple biological processes are analysed in unison (Chari et al., 2010). Gene co-expression network analysis may help to facilitate this goal by providing an endpoint for the quantification of such processes. For example, Parikshak et al. (2013) found that by intersecting co-expression modules with GWAS loci, they were able to identify ASD genes that tightly coalesced with modules implicated in distinct biological functions during human cortical development, including early transcriptional regulation and synaptic development. Furthermore, when modules of a network are combined with epigenetic information, we can substantially enrich our understanding of the epigenetic interplay between both genes and the environment in psychiatric disorders (Gibney and Nolan, 2010). As such, integrated transcriptome and methylome data derived from peripheral blood was able to identify 43 risk genes that discriminated youth patients and high-risk for bipolar disorder patients from controls (Fries et al., 2017), further demonstrating the utility of integrative analyses in the identification of biomarkers for disease risk.

Combining more than two layers of biological data is a largely unexplored avenue in psychiatric research (Bersanelli et al., 2016). However, Ciuculete et al. (2017) investigated the interplay between 37 psychiatric-related genetic risk variants as well as shifts in both methylation and mRNA levels in 223 adolescents distinguished as being at risk for the development of psychiatric disorders. Using this approach, the authors were able to detect five SNPs (in *HCRTR1, GAD1, HADC3* and *FKBP5*), which were associated with eight CpG sites. Three of these CpG sites, cg01089319 (*GAD1*), cg01089249 (*GAD1*) and cg24137543 (*DIAPH1*), manifest in significant gene expression

changes and overlap with active regulatory regions in chromatin states of brain tissues. Although these findings are preliminary, further studies employing multi-stage integrative analysis may help to advance the field and provide novel insights on how genetic variants may modulate risks for the development of specific psychiatric diseases (Ritchie et al., 2015).

Shared molecular mechanisms between disorders: implications for treatment response

There is growing understanding that many psychiatric diseases share underlying biological mechanisms. The cross-disorder group from the Psychiatric Genomics Consortium (PGC) has provided empirical evidence for the shared genetic etiology of five psychiatric disorders, including, schizophrenia, bipolar disorder, major depressive disorder (MDD), autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD) (Lee et al., 2013). Recently, (Lotan et al., 2014) curated all GWAS findings for ADHD, anxiety disorders, ASD, bipolar disorder, MDD, and schizophrenia, finding that 22% of identified genes overlapped across two or more disorders. An overlap in underlying biology between different conditions was also observed at the level of gene expression. For example, using microarrays, Iwamoto et al. (2004) compared gene expression profiles across different brain regions in bipolar disorder, MDD and schizophrenia. What they found is that although these mental disorders were molecularly distinctive, there was a relatively large overlap of the gene expression profiles in all three disorders. Using RNA-seq, further support was provided for these findings by Darby et al. (2016), finding that hippocampus and orbitofrontal cortex transcriptomes were consistent across diagnostic groups. These fundings may also help to explain why the same medication exerts its effects on patients diagnosed with different disorders.

Although we have outlined many studies that have successfully exploited co-expression network approaches in psychiatric research, this methodology is yet to be fully

embraced by the field. New methods and algorithms for gene expression analyses are growing. Including the use of co-expression networks in conjunction with machine learning methodologies (Lareau et al., 2015). In the following section the role of machine learning in both genetics and gene expression for developing predictive models for diagnostic purposes as well as for treatment response will be further explored.

Future Methods: Machine Learning

Machine learning (ML) exists at the intersection of computer science, mathematics, and statistics and is proving to be one such methodology that can handle the modelling of large, complex systems datasets (Iniesta et al., 2016). Given the brevity of this chapter, only a very brief explanation will be provided (see Iniesta et al. (2016) for a more extensive review). Firstly, supervised ML algorithms learn from data and improve their performance through experience. This is achieved through learning a set of features (SNPs, gene expression data) that relate to a label. The label can be either binary (response/non-response) or multiclass (low-response/moderate-response/high-response). Supervised ML can also be used for regression, where a continuous outcome such as psychiatric illness severity can be predicted (Hastie et al., 2009). In contrast, unsupervised ML algorithms identify clusters in data derived from a distance metric (such as Euclidean distance) that potentially contain unique characteristics related to an outcome of interest (Ghahramani, 2004). In this section, we will focus on supervised ML models.

Gene selection and complex trait prediction

There are two main problems that ML may help to resolve in both gene expression and genome-wide association studies. Firstly, given the large p small n problem ($p \gg n$), a methodology is needed that can not only reduce the dimensionality of a genetic feature space, but do so in a way that can lead to a clinically meaningful prediction of a complex

trait. For example, using a patient's gene expression data, can we predict whether or not a patient will respond to drug A) or drug B)? In differential expression analysis (DEA), univariate statistical tests are often performed to discover quantitative changes in expression levels between case and control phenotypes (Gupta et al., 2012). However, according to Okser et al. (2014), regularized ML models have demonstrated an improved ability to select SNPs and differentially expressed genes that are most predictive of complex traits. In ML, a process called regularization penalises a models' complexity, thus, enabling prediction in individuals outside of a training data set. A beneficial characteristic of this approach is that they search out the most predictive combinations of variants, rather than just individually predictive variants like in univariate statistical tests. It is therefore not surprising that variants that do not attain genome-wide significance in univariate tests, often contribute to the predictive capacity of multi-locus ML models (Abraham et al., 2013, Evans et al., 2009). Furthermore, evidence is beginning to suggest that genetic markers with highly significant and replicated odds ratios derived from genome wide association studies (GWAS) may actually be poor classifiers of disease (Jakobsdottir et al., 2009). This is not to suggest that findings from GWAS studies are not useful; they provide valuable information for establishing etiological hypotheses. What it may suggest is that regularized ML models may be more beneficial for the derivation of clinically translatable pathways and variants that evade detection in classical statistical methods when $p \gg n$. Therefore, the use of regularized ML models may lead to greater prediction of disease and overall utility in translational psychiatry.

Support for this hypothesis has been demonstrated by (Wei et al., 2013) in their ML based prediction of Crohn's disease. Using regularized (lasso) logistic regression, they trained a prediction model whist also minimising a feature space of 10,799 SNPs. Their final model contained 573 SNPs and obtained an area under the curve (AUC) of 0.86. They then compared this model to a traditional log odds model. For dimensionality reduction,

they estimated association significance and odds ratios for each variant by using two folds of data. They took all variants where $p < 0.01$ and pruned correlated variants by setting the linkage disequilibrium threshold to $r^2 \frac{1}{4}\ 0.8$. The final feature space contained 15,158 SNPs and obtained an AUC of 0.73, a score 13% lower than the penalized logistic regression model whilst containing a feature space that was more than 25 times the size. Thus, achieving both sub-optimal predictive performance and decreased computational tractability relative to the regularized ML model.

Applications for gene expression

The use of ML methodologies for both dimensionality reduction and prediction of complex traits in gene expression studies is less prevalent, however, some support for this methodology has been attained. For example, (Tan and Gilbert, 2003) trained a decision tree classifier with probes from 6,817 genes to predict cancerous colon tumours, attaining an accuracy score of 95% to discern between each type of tumour. Furthermore, using bagged decision trees, they were also able to differentiate between two different types of lung cancer with 93% accuracy. Whether such findings are possible in complex psychiatric traits such as major depression is questionable, yet the methodology is theoretically applicable; and as demonstrated; appears to outperform univariate statistical tests and classical methods for both dimensionality reduction and prediction of complex traits when $p \gg n$. In support of this, preliminary evidence suggests that the use of an ML based methodologies in complex psychiatric trait prediction may be possible, yet much larger studies are still needed. For example, using a prospective design, Guilloux et al. (2014) collected blood samples from a discovery cohort of 34 adult MDD patients with co-occurring anxiety and 33 matched non-depressed controls. Data was collected at baseline and after 12 weeks of combined citalopram and psychotherapy treatment. Using linear support vector machines trained on gene-expression data from 13 genes, they were

able to predict remission/non-remission of MDD with a cross validated corrected accuracy score of 76.2% (sensitivity = 86.1%, specificity = 59.3%). Much larger studies are still warranted across a range of complex traits and treatments; however, the initial proof of concept is encouraging.

Conclusions

Studying interactions between genes in relation to psychiatric phenotype using co-expression network analysis is a promising complementary approach to better understand molecular mechanisms of psychiatric disorders, which may lead to developing clinically translatable diagnostic and treatment response biomarkers. However, further research needs to also utilise machine learning algorithms to quantify diagnosis and treatment response prediction. We believe that these methods, combined, can help to advance a rapidly developing field of personalised psychiatry.

Appendix A

Supplementary materials for the manuscript

Co-expression network analysis of peripheral blood transcriptome identifies dysregulated protein processing in endoplasmic reticulum and immune response in recurrent MDD in older adults

Content

Appendix A text

1. Demographics and case definition for sMAS and OATS cohorts
    1.1. Sydney Memory and Aging Study (sMAS)
    1.2. Older Australians Twins Study (OATS)
2. Details on data generation and analyses
    2.1. Purification of intracellular RNA
    2.2. Whole-genome gene expression direct hybridization
    2.3. Data quality control (QC)
    2.4. Data pre-processing
    2.5. Weighed gene co-expression network analysis (WGCNA)
    2.6. Type-specific markers of expression profiles in recurrent MDD-associated modules
    2.7. Functional analyses
    2.8. Replication (OATS cohort)
    2.9. Significance of the overlap between sMAS and OATS findings

Appendix A Tables

Table S1.1. Statistical tests of independence for Age, Sex and BMI between cases and controls for four MDD patient subgroups in MAS cohort

Table S1.2. Number of modules identified and the module sizes

Table S1.3. Number of probes associated with recurrent MDD at FDR 0.2

Table S1.4. Number of cell type-specific markers found in recurrent MDD-related modules

Table S1.5. Module preservation in sMAS and OATS datasets quantified by Z summary score

Table S1.6. Replication of MAS findings in OATS

Table S1.7. Nine common genes identified across three modules in the discovery (MAS) and replication (OATS) cohorts as associated with recurrent MDD in older adults

Table S1.8. Statistical significance of the overlap between recurrent MDD-related probes in the discovery (sMAS) and replication (OATS) cohorts

Table S1.9. List of 24 genes belonging to recurrent MDD-associated modules that previously were found as differentially expressed between MDD cases and controls

Appendix A Figures

Figure S1.1. Housekeeping genes performance

Figure S1.2. System background noise

Figure S1.3. Array hybridization control

Figure S1.4. Signal generation control

Figure S1.5. Intensity boxplots (mean and SD) before pre-processing and after pre-processing (background correction, transformation, normalization)

Figure S1.6. Density plot of intensity before pre-processing and after pre-processing

Figure S1.7. Filtering by coefficient of variation. Red line represents a threshold of 0.01.

Figure S1.8. Density plot of intensity of pre-processed and filtered MAS data

Figure S1.9. Scale independence and mean connectivity

Figure S1.10. Clustering dendrogram of genes

Figure S1.11. Visualisation of relative proportions of recurrent MDD-associated modules' genes found in cell type-specific gene expression signatures

Figure S1.12. General Network properties defined as correlations between ranked expression and ranked connectivity for 10654 genes expressed in both MAS and OATS data sets

Supplementary References

1. Demographics and case definition for the MAS and OATS

1.1. Sydney Memory and Aging Study (MAS)

The Sydney Memory and Aging Study (MAS) was initiated in 2005 to examine the clinical characteristics and prevalence of mild cognitive impairment (MCI) and related syndromes, including depression, and to determine the rate of change over time. At the baseline assessment from 2005 to 2007, 1037 non-demented individuals aged 70-90 were recruited from two areas of Sydney, following a random approach to 8914 individuals on the electoral roll. They underwent detailed neuropsychological assessment and donated a blood sample for clinical chemistry, proteomics and genomics. Participants were excluded if they had a previous diagnosis of dementia, psychotic symptoms or a diagnosis of schizophrenia or bipolar disorder, multiple sclerosis, motor neuron disease, developmental disability, progressive malignancy (active cancer or receiving treatment for cancer, other than prostate – non-metastases, and skin cancer), or if they had medical or psychological conditions that may have prevented them from completing assessments. Participants were excluded if they had a Mini-mental Statement Examination (MMSE) (Folstein et al., 1975) score of <24 adjusted for age, education and non-English speaking background at study entry, or if they received a diagnosis of dementia after a comprehensive assessment. Participants were followed up with brief telephone reviews annually and detailed assessments biannually. More details on study design are published somewhere else (Sachdev et al., 2010).
For the purposes of current work, we utilised the MAS as an exploratory cohort. Participants who had their venous blood collected for transcriptome analysis in Wave 4 in MAS (N=521) were included in further analyses. The demographic characteristics of the sample are presented in Table S1. The mean age of the sample was 77.84 years at Wave 1, and 83.73 years at Wave 4. The sex ratio was F 255/M 266. Participants had a mean education of 11.8 years. Just over 99% of the participants were of Caucasian origin (Table S1.1).

Table S1.1. Statistical tests of independence for Age, Sex and BMI between cases and

controls for four MDD patient subgroups in MAS cohort

	Age			Sex			BMI		
MDD patient subgroup	**MDD cases mean±SD**	**Controls mean±SD**	**t-test**	**Number of MDD cases M / F**	**Number of controls M / F**	**X^2 test**	**MDD cases mean±SD**	**Controls mean±SD**	**t-test**
LD	84.2±4.9	83.6±4.4	t(519)=1.02, p=0.3	33 / 47	222 / 219	X^2(3, N=521)=2.21, p=0.13	27.2±4.3	26.7±4.2	t(519)=0.99, p=0.3
SD	83.7±4.6	83.7±4.6	t(519)=0.9, p=0.3	21 / 35	234 / 231	X^2(3, N=521)=3.29, p=0.07	27.3±4.4	26.8±4.2	t(519)=0.8, p=0.4
CD	84.8±4.9	83.6±4.4	t(519)=1.7, p=0.07	20 / 31	236 / 234	X^2(3, N=521)=2.23, p=0.13	27.6±4.3	26.7±4.2	t(519)=1.4, p=0.2
RD	83.7±4.5	83.8±4.0	t(519)=0.1, p=0.9	14 / 13	241 / 253	X^2(3, N=521)=0.09, p=0.8	27.4±4.4	26.8±4.2	t(519)=0.7, p=0.5

LD – Lifetime diagnosis of MDD, SD – Single episode of MDD, CD – Current MDD, RD – Recurrent MDD

Depression in MAS was assessed by a battery of well-validated self-report and interview-based measures of depression. This buttery included: Geriatric Depression Scale (GDS-15), Neuropsychiatric Inventory (NPI) (depression sub-scale), Mini International Neuropsychiatric Interview (M.I.N.I.), and Patient Health Questionnaire (PHQ-9). Given that the measures availability differed across the waves and the data in some of the measures were sparse, we argue that integration of all available data in time point diagnosis of depression at each wave is a meaningful approach to overcome these limitations. To integrate all available data on different measures of depression across the waves we created a composite variables depression for each wave. These composite scores represent a summative diagnosis of depression using all available information about the clinical representation of depression for 521 individuals. The step-by-step process of creating the composite variables of depression is detailed below.

The process of creating composite variables of depression consisted of three steps:

1) Creating unified dichotomous variables of depression for each measure in each wave.

2) Creating a composite diagnosis of depression for each wave using "minimum by two" approach.
3) Creating a lifetime MDD, single episode, current, recurrent MDD variables of depression

Step 1. Creating unified dichotomous variables of depression for each measure in each wave.

Measures included:

Geriatric Depression Scale (GDS-15)

The Geriatric Depression Scale (GDS-15) is a 15-item self-report assessment used to identify depression in the elderly (Yesavage et al., 1982b). In the MAS we used the GDS-15 with item 9 as described in Brink (1982, here item 12) instead of the more common one. In order to determine whether this would affect the classification of depression, a statistical analysis was carried out on two data sets from the same sample, one with the GDS-15 containing the version used in MAS, and the other with the more common version. Using the same cut-point for depression (score ≥ 6) the results showed a high level of agreement between the two classifications (kappa = .931) with a total of only 4 false positives, and no false negatives, out of a sample of 354.

To derive a dichotomous variable of GDS scoring we applied commonly used a cut-off of 6; participants scored ≥6 were assigned to 1 (depressed by GDS-15) and 0 (not depressed) if scored <6.

Neuropsychiatric Inventory (NPI). Depression sub-scales.

The Neuropsychiatric Inventory (NPI) was developed by Cummings et al. (1994b) to assess dementia-related behavioural symptoms. The ratings of the NPI produce one score per behavioural domain. This score reflects the degree of disturbance of a particular domain. In this study, we used NPI sub-scales of depression to screen for depressive behaviours in our sample. Although using an individual NPI sub-scales have been subjected to further testing for its clinical utility (Lai, 2014), depression sub-scale showed moderate correlation with well-established Montgomery-Asberg Depression Rating Scale (MADRS) and Cornell Scale for Depression in Dementia (CSDD) depression instruments (MADRS: r = 0.70; CSDD: r = 0.62) (Leontjevas et al., 2009). This indicates that the NPI depression sub-scales are suitable for research purposes.

A participant was assigned to 1 (depressed by NPI depression sub-scale) if a diagnosis was confirmed and to 0, if the diagnosis of depression was not confirmed.

Patient Health Questionnaire (PHQ-9)

The Patient Health Questionnaire (PHQ) is a self-administered version of the PRIME-MD diagnostic instrument for common mental disorders. The PHQ-9 is the depression module to assess depression severity; it scores each of the 9 DSM-IV criteria as “0” (not at all) to “3” (nearly every day). A PHQ-9 score ≥10 (moderate to severe depression) was shown to have a sensitivity of 88% and a specificity of 88% (Kroenke et al., 2001) for major depression. In this study, we used the cut-off of ≥10 to identify participants with depression.

The Mini International Neuropsychiatric Interview (M.I.N.I.)

The M.I.N.I. is a widely used psychiatric structured diagnostic interview instrument for the major Axis I psychiatric disorders in DSM-IV and ICD-10. The MINI is divided into modules identified by letters corresponding to diagnostic categories. Each diagnostic module consists of screening questions corresponding to the main criteria of the disorder. All questions are rated and clinical judgment used in coding the responses. At the end of each module, clinician indicate whether diagnostic criteria are met. In this study, to create a dichotomous variable of depression by M.I.N.I. we used the final diagnosis in the category of depression made by clinicians.

Step 2. Creating a composite diagnosis of depression for each wave using “minimum by two” approach**.**

In the second step, we summarised by how many measures of depression participants were diagnosed by interview-based or/and self-report measures at each wave. This summative measure reflected the level of diagnostic consensus across the measures within each Wave. Given that the measures availability differed across the waves and the data in some of the measures were sparse, we argued that integration of all available data in time point diagnosis of depression at each wave is a meaningful approach to overcome these limitations. To make use of all measures available, a “minimum by two” approach for diagnosing MDD was defined. Accordingly, MDD cases were defined as participants who were identified as clinically depressed by at least two of the above depression measures, according to DSM-IV criteria at each Wave.

Step 3. Creating lifetime MDD, single episode, current, recurrent MDD variables of depression

In the third step, we created composite variables of MDD (lifetime, single episode, current, and recurrent) across the waves described in the main text and Table 1.

The 3 steps approach applied to MAS phenotype data of depression resulted in 13.3% of

lifetime depression prevalence, which is in line with current estimations of global prevalence of depression in older adults (Fiske et al., 2009).

1.2. Older Australians Twins Study (OATS)

The Older Australian Twins Study is a longitudinal study investigating healthy brain ageing in older twins (65+ years). Healthy ageing is characterised by low levels of disability, high cognitive and functional capacity, and an active engagement in life. The most important ingredient of healthy ageing is a healthy brain, bereft of age-related diseases and dysfunction, including depression. OATS commenced in New South Wales in January 2007, in Queensland in December 2007, and in Victoria in February 2008. Since the OATS study started we have followed our twin volunteers up every two years to check on their psychological and physical health. Participants undergo rigorous medical and cognitive function tests, with many participants also providing blood samples and having a magnetic resonance imaging (MRI) scan of their brain. OATS assessed 623 participants at baseline, 450 at the 2-year follow up, and 390 completed their 4-year follow-up (Sachdev et al., 2012).

For the purposes of current work, we utilised the OATS as a replication cohort.

186 unrelated OATS participants, which had their venous blood collected for transcriptome analysis in Wave 3 were randomly selected (one twin from each twin pair) for further analyses.

Depression in OATS was assessed by self-report (GDS-15 and PHQ-9) interview-based measures (NPI, depression sub-scale) and SCID. All depression assessment measures of depression were consistent between MAS and OATS, except M.I.N.I./SCID. While in MAS was used M.I.N.I. OATS utilised SCID.

The Structured Clinical Interview for DSM (SCID) is a semi-structured interview guide for making the major DSM diagnoses. It was administered by a clinician or trained mental health professional.

Using the three steps methodology developed for MAS, we identified the subgroup of recurrently depressed individuals in OATS cohort (N=7), which was used to conduct replication analysis.

2. Details on data generation and analyses

2.1. Purification of intracellular RNA

To extract RNA form the whole blood we used the PAXgene Blood RNA System (PreAnalytiX, QIAGEN), which consists of a blood collection tube (PAXgene Blood RNA Tubes) and nucleic acid purification kit (PAXgene Blood RNA Kit). It is intended for the collection, storage, and transport of blood and stabilization of intracellular RNA in a closed tube and subsequent isolation and purification of intracellular RNA from whole blood.

The whole blood of 536 MAS participants at wave 4 was collected into PAXgene Blood RNA Tubes for stabilization further processing after overnight fasting between 7 and 9 am.

The total RNA was successfully extracted from whole blood collected from 521 samples using PAXgene Blood RNA Kits following manufacturer's protocol for manual purification. In short, 2.5 ml of whole blood was centrifuged to pellet nucleic acids in the PAXgene Tube. The pellet was washed and resuspended. The resuspended pellet was incubated in optimized buffers together with proteinase K to bring about protein digestion. After homogenising the cell lysate and removing residual cell debris, the supernatant of the flow-through fraction was transferred to a fresh microcentrifuge tube. Ethanol was added to adjust binding conditions, and the lysate was applied to a PAXgene spin column. During a brief centrifugation, RNA was selectively bound to the PAXgebe silica membrane as contaminants pass through. Remaining contaminants were removed in several wash steps. Between the first and second wash steps, the membrane was treated with DNase I to remove trace amounts of bound DNA. After the wash steps, RNA was eluted in elution buffer and heat-denatured.

2.2. Whole-genome gene expression direct hybridization

The gene expression data were generated using the Illumina Whole-Genome Gene Expression Direct Hybridization Assay System *IlluminaHT-12 v4*. The *HumanHT-12 v4* Expression BeadChip content provides genome-wide transcriptional coverage of well-characterized genes, gene candidates, and splice variants, delivering high-throughput processing of 12 samples per BeadChip. Each array on the *HumanHT-12 v4* Expression BeadChip targets more than 47,000 probes derived from the National Center for Biotechnology Information Reference Sequence (NCBI) RefSeq Release 38 (November 7, 2009) and other sources. The gene expression data was generated in the Adelaide

Microarray Centre (SA Pathology, Adelaide, Australia) by qualified personnel following manufacturer's protocols. In short, unlabelled total RNA extracted from the whole blood samples were biotin-labelled and amplified producing a pool of cRNA corresponding to the polyadenylated (mRNA) fraction. The labelled RNA strand was then hybridized to the bead on the BeadChip containing the complementary gene-specific sequence. After the overnight hybridization, the BeadChips were removed and then washed. To detect a signal, Cy3-SA was introduced to bind to analytical probes that have been hybridized to the BeadChip. This allowed for differential detection of signals when BeadChips were scanned. The fluorescence intensity was scanned at each addressed bead location using the scanner Illumina HiScan. The intensity of the signal should correspond to the quality of the respective transcript in the original sample. The raw intensity images were analysed using Illumina's Genome Studio Gene expression Module.

2.3. Data quality control (QC)

Quality control was performed in R environment using limma package.

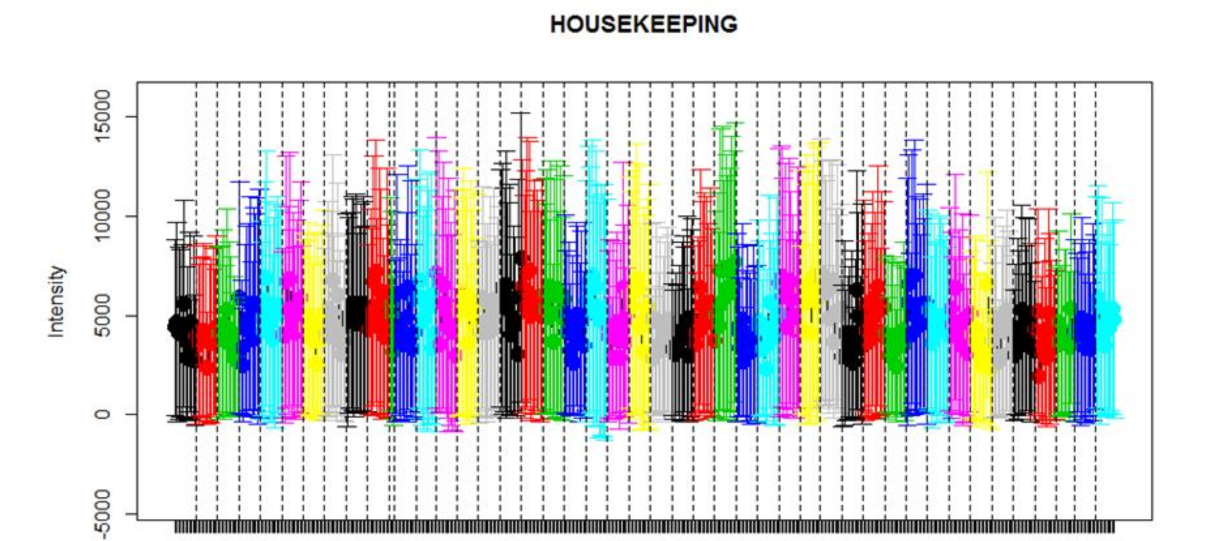


Figure S1.1. Housekeeping genes performance

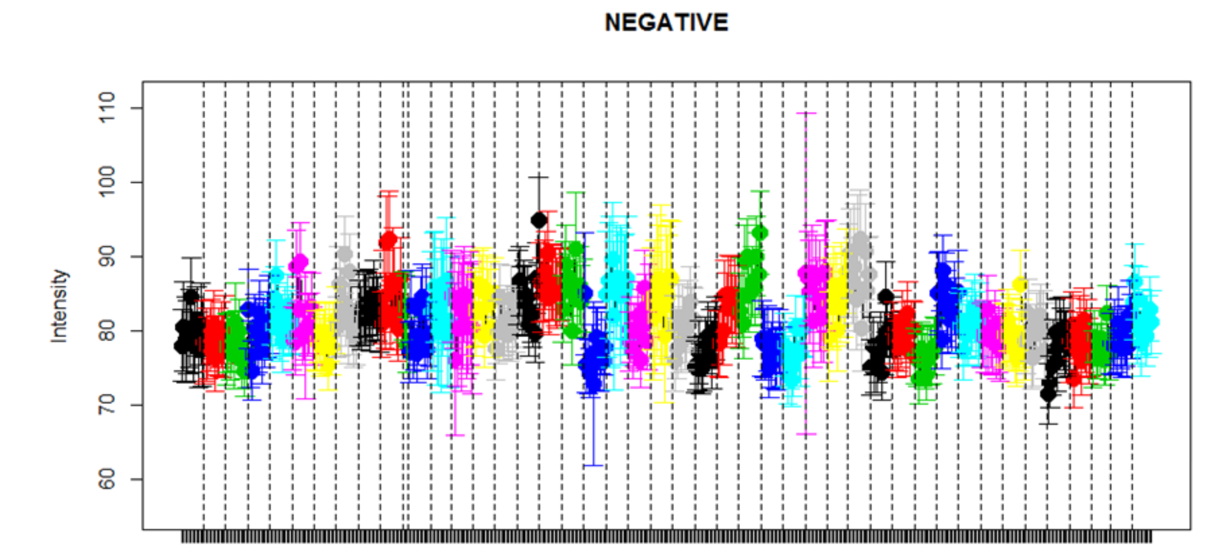


Figure S1.2. System background noise

The probes in this control category correspond to Cy3-labeled oligonucleotides. Following successful hybridization, they produce a signal independent of both RNA quality and success of the sample prep reactions.

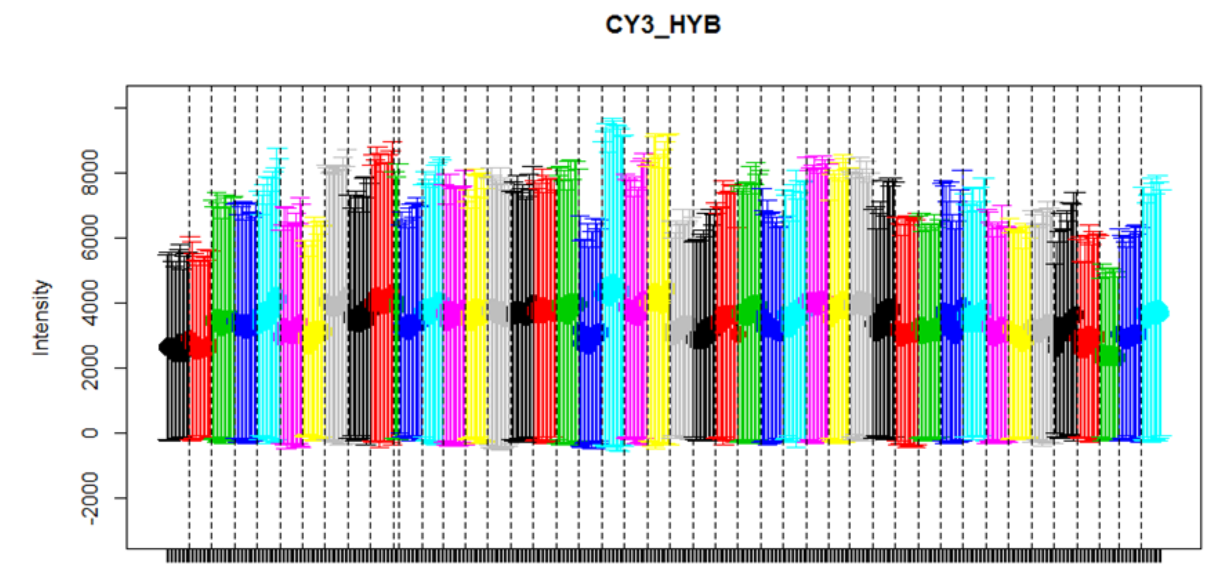


Figure S1.3. Array hybridization control

Signal generation control consists of two probes with complementary biotin-tagged

oligonucleotides. Successful secondary staining is indicated by a positive hybridization signal from these probes.

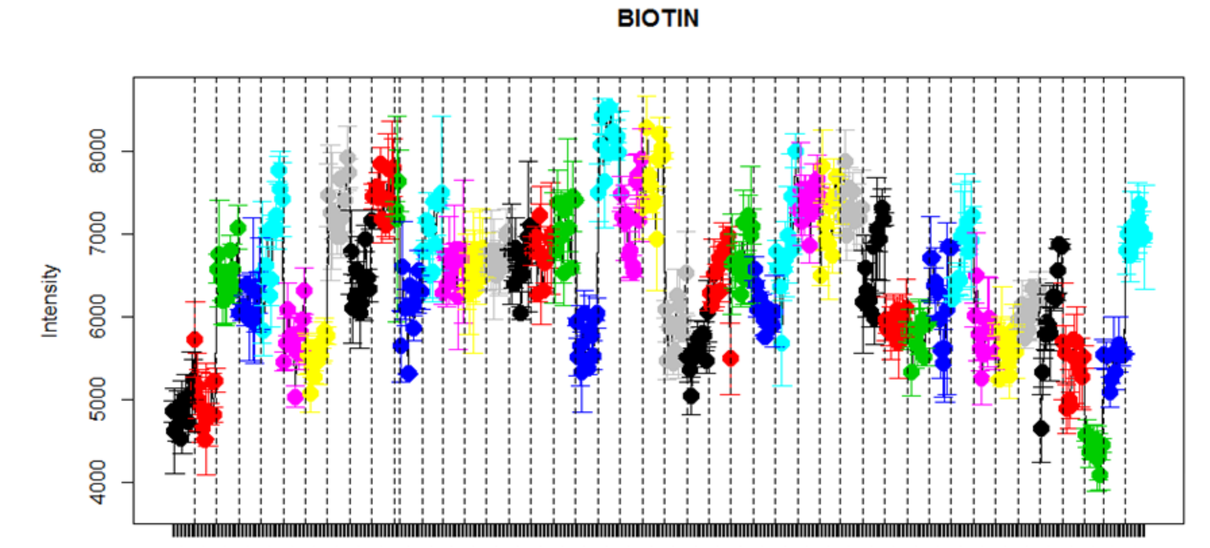


Figure S1.4. Signal generation control

2.4. Data pre-processing

Background correction. Maximum Likelihood Estimation (MLE) algorithm

To control for background noise, which is inherited in gene expression experiments, Illumina allocated more than 1000 control bead types to each array, which are not associated with gene-specific probes, i.e. have no corresponding targets in the genomes. Therefore, the control beads are not expected to hybridize to any genes in the RNA samples. They serve as a comprehensive measurement of background, representing the imaging system background noise as well as any signal resulting from non-specific binding of dye or cross-hybridization in an experiment (Xie et al., 2009). The signals and signal standard deviation of these probes were used to establish gene expression detection limits (GenomeStudio). A model-based background correction implemented in R package MBCB allows incorporating information from these negative controls, which removes the non-specific signal from total signal and improves the efficiency of background correction, leading therefore to more precise determination of gene expression and better biological interpretation of the data (Allen et al., 2009). The maximum likelihood estimation (MLE) method showed the best estimation efficiency over

the other methods available in MBCB package (non-parametric (NP), robust multi-array average (RMA), Bayesian (B)) (Xie et al., 2009).

Variance-stabilising transformation (VST)

Variance stabilization is critical for subsequent statistical interference from microarray data. Illumina microarray platform provides a large number of technical replicates on each array (over 30 spatially distributed at random locations beads per probe). We performed the variance-stabilizing transformation (VST) method takes advantage of these technical replicates to stabilize the variance better and more efficiently (Lin et al., 2008). The VST requires the information of the standard deviation, however, the BeadStudio output file provides the standard error of the mean of the bead intensities corresponding to the same probe. Therefore, correction is required. The corrected value will be x * sqrt (N), where x is the standard error of the mean and N is the number of beads corresponding to the probe. VST can be viewed as a generalized log2 transformation, fine-tuned for the noise characteristics of each array.

Normalization (quantile)

When preparing microarray data for downstream analyses, it is important to remove sources of variation between arrays of non-biological origin. Normalization is a process for reducing this variation. There are different methods for data normalization available; quantile normalization is one of the most commonly used methods due to its robustness and performance speed (Bolstad et al., 2003).

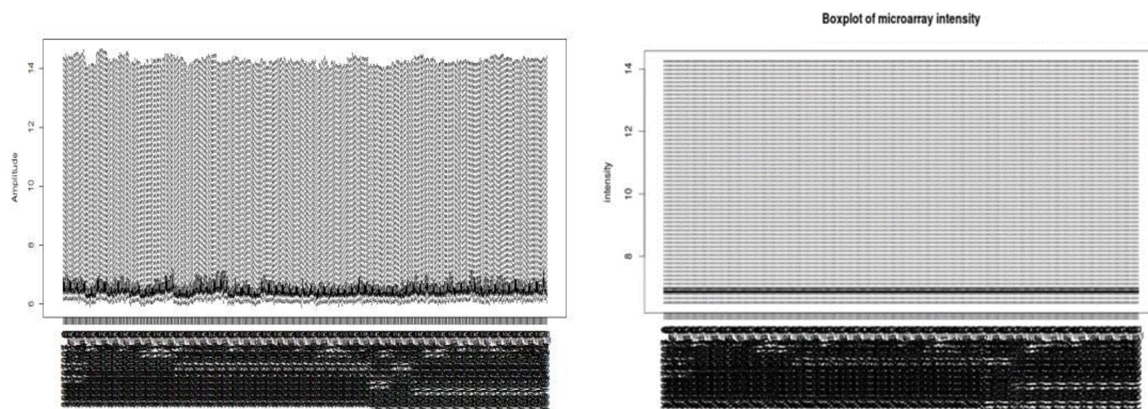


Figure S1.5. Intensity boxplots (mean and SD) before pre-processing and after pre-processing (background correction, transformation, normalization)

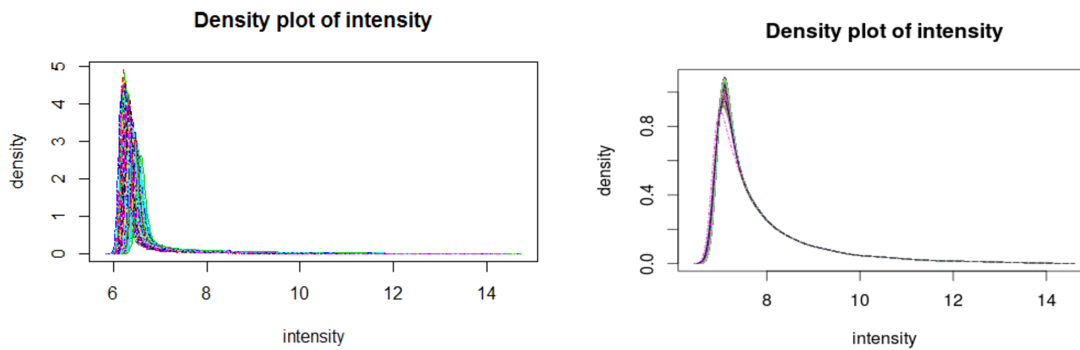


Figure S1.6. Density plot of intensity before pre-processing and after pre-processing

Filtering by detection value

After filtering probes by a stringent criteria of detection p-value <0.01 in ≥50% of samples we had a dataset of 12,852 individual probes.

Filtering by coefficient of variation

For the purpose of the gene co-expression analysis we are interested in the genes the expression of which vary across the samples, since low-expressed or non-varying genes usually represent noise. To filter out low-varying probes we used coefficient of variation. The coefficient of variation shows the extent of variability in relation to the mean of the sample and is defined as the ratio of the standard deviation to the mean. CV was computed for each probe across all samples. The threshold of 0.01 (red line) is used to filter out the probes with low variation. 1834 genes were filtered out giving us a final data set of 11, 018 probes corresponding to 9041 genes.

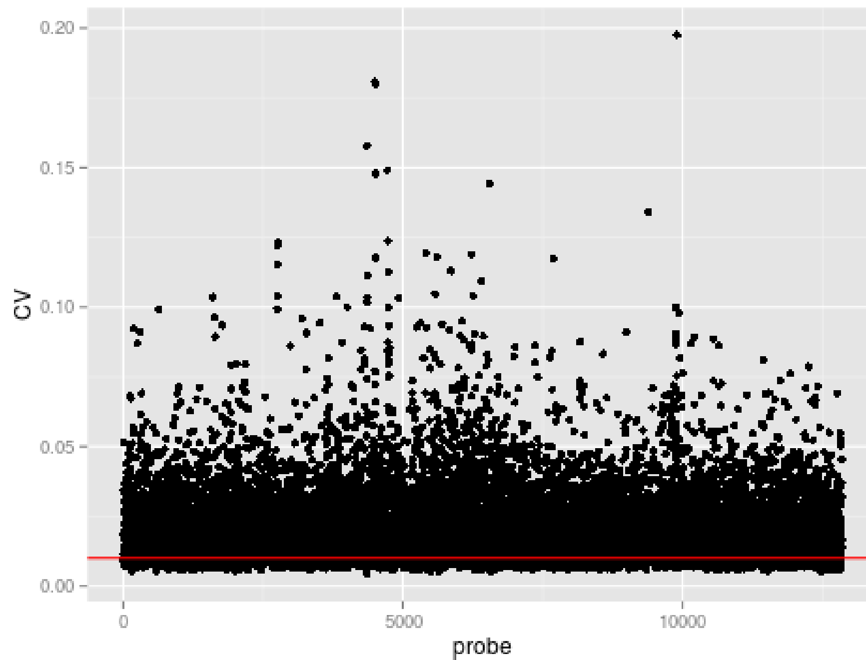


Figure S1.7. Filtering by coefficient of variation. Red line represents a threshold of 0.01.

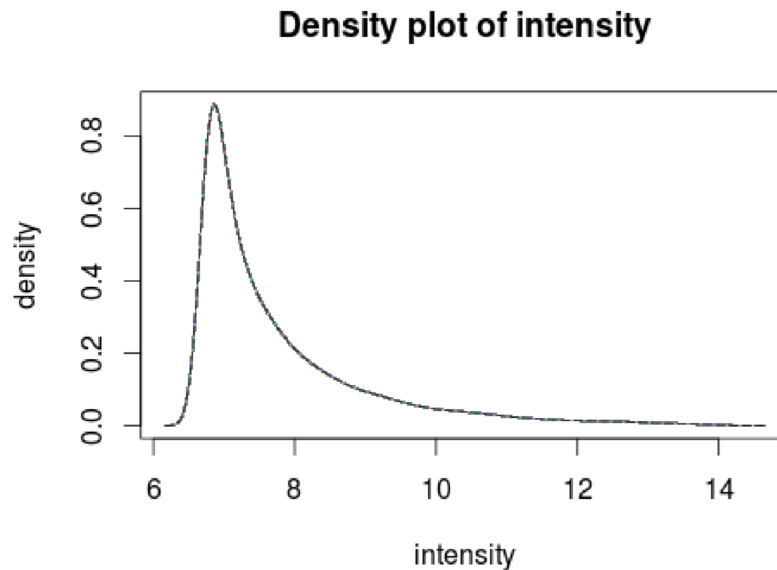


Figure S1.8. Density plot of intensity of pre-processed and filtered MAS data

Data adjustment for technical covariates

Data was adjusted for batch and RINs effects using Empirical bayes-moderated linear regression implemented in *empiricalBayesLM* function (WGCNA package).

To estimate latent non-biological variables we used *sva* function (sva R package) (Leek et al., 2012). The number of significant surrogate variables was 0 (in both, MAS and OATS datasets), therefore there was no need to adjust the data for these factors.

2.5. Weighed gene co-expression network analysis (WGCNA)

WGCNA was performed using the WGCNA R package, as previously described (Langfelder and Horvath, 2008). In short, this method selects the threshold for constructing the network based on the scale-free topology of gene co-expression networks. Using biweight mid-correlation, which is a median-based rather than mean-based robust to outliers alternative of Pearson's correlation implemented in *bicor* function (Langfelder and Horvath, 2012), we computed the networks for several thresholds and selected the threshold β=6, which led to a network with scale-free topology (Figure S1.9).

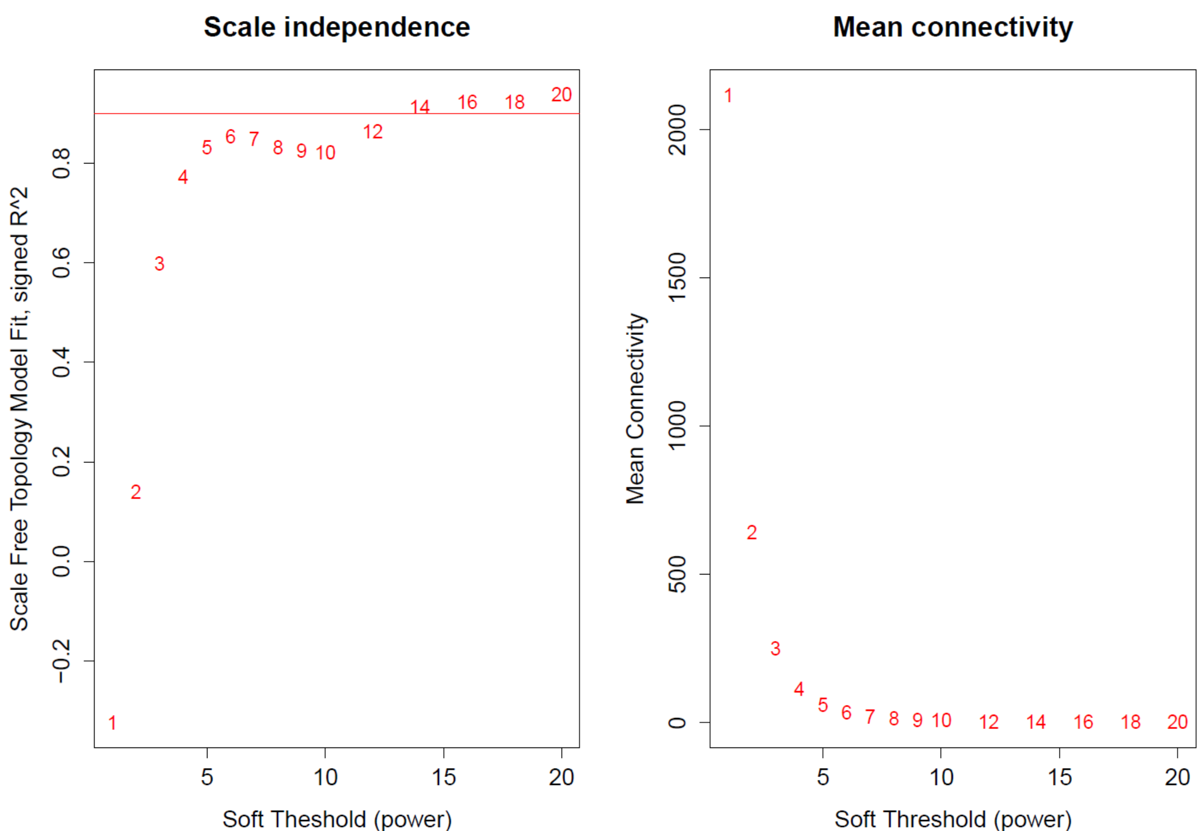


Figure S1.9. Scale independence and mean connectivity**.**

Analysis of network topology for various soft-thresholding powers. The left panel shows the scale-free fit index (y-axis) as a function of the soft-thresholding power (x-axis). The right panel displays the mean connectivity (degree, y-axis) as a function of the soft-thresholding power (x-axis).

The network consisted of 24 modules of functionally related genes on the basis of their

expression patterns. The modules were labelled 1 through 24 in order of descending size, with sizes ranging from 2845 to 34 genes. The label 0 is reserved for genes outside of all modules (Table S1.2).

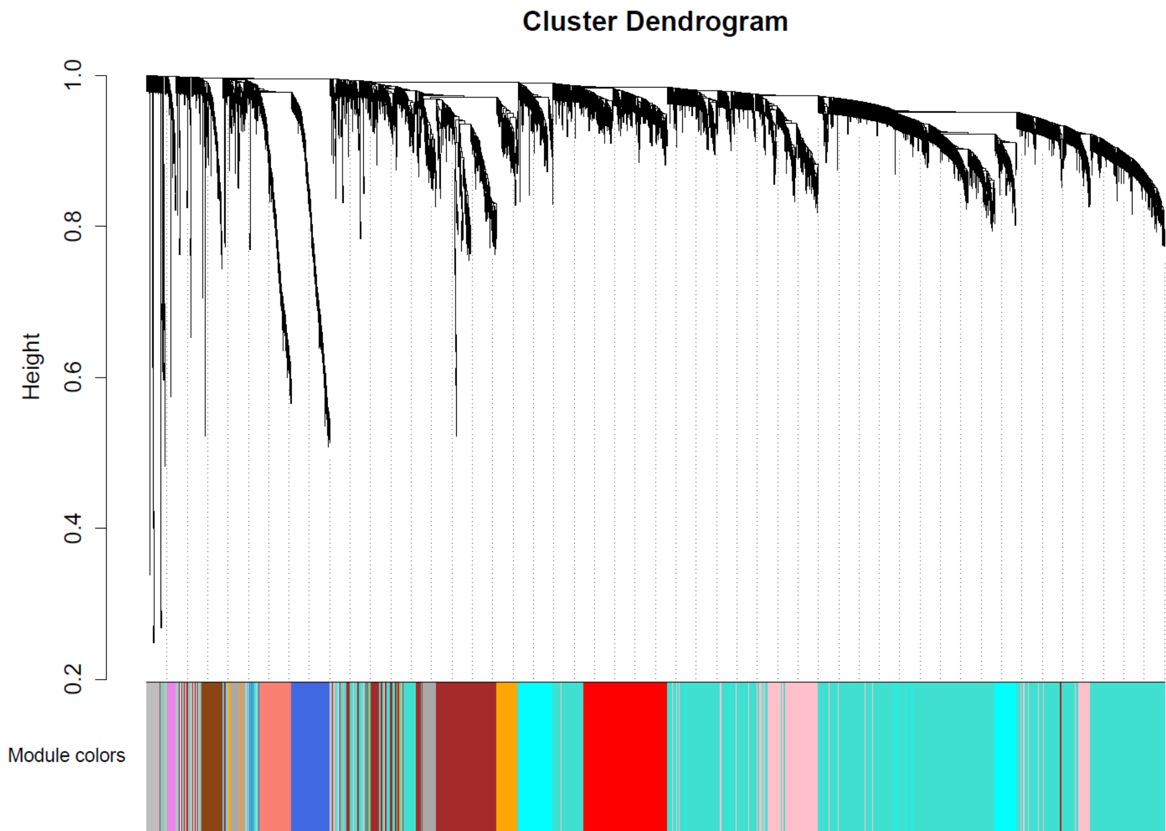


Figure S1.10. Clustering dendrogram of genes, with dissimilarity based on topological overlap, together with assigned module colours.

Table S1.2. Number of modules identified and the module sizes.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
926	2845	925	601	582	507	434	410	409	398	368	358	315	252	232	226	185	182	160	139	129	120	108	91	82	34

To quantify the relationship between MEs and binary outcome variable (MDD status, 1 or 0) we used *bicor* function with the parameters *robustX* and *robustY* set to FALSE, with

reverses robust biweight midcorrelation to standard Pearson's correlation (Langfelder and Horvath, 2012).Correction for multiple testing

To correct obtained p-values for GS measure for transcriptome-wide multiple testing we used FDR at relaxed level of 0.2. The reason behind choosing this liberal threshold of FDT at 0.2 is based on previous literature suggesting that depression is a heterogeneous disorder with multiple genes of a small effect. Given the limited sample size of MDD cases, this study is underpowered to detect MDD-associated genes with a better precision. However, we believe that we were able to capture depression-relevant biological signal that can be validated in the future studies with larger sample sizes.

There were 1,656 probes nominally associated with recurrent MDD ($p < 0.05$); after transcriptome-wide correction for multiple comparisons using Benjamini-Hochberg procedure, we identified 923 recurrent MDD-associated probes at FDR = 0.2; 82% of those (761 probes) belonged to the modules the MEs of which showed association with recurrent MDD. 162 probes belonged to other modules of the co-expression network.

Table S1.3. Number of probes associated with recurrent MDD at FDR 0.2. DOWN represents downregulated genes, UP – upregulated.

Module name	Total N of probes	DOWN, N (%)	UP, N (%)
Turquoise	614	565 (92%)	49 (8%)
Tan	56	0	56 (100%)
Black	58	0	58 (100%)
Lightcyan	33	2 (6%)	31 94%)

There were no probes associated with recurrent MDD in the replication study (OATS) after correction for multiple testing at FDR 0.2.

2.6. Type-specific markers of expression profiles in recurrent MDD-associated modules

Whole blood is a complex tissue consisting of numerous cell types. To assess whether specific cell type/s driving the expression signature we found as relevant for recurrent MDD, we obtained known gene expression signatures for six major immune cell types and contrasted these to the modules of interest derived from our data. To define cell type-specific gene expression signatures, we used the most recent in-silico estimations calculated from multiple data sources. Authors employed a curve fitting approach for linear comparison of cell types on harmonized 1822 pure human cell type transcriptomes from various sources, and introduced a novel spillover compensation technique for separating them (Aran et al., 2017). We combined cell-type specific markers obtained from 5 data sources (FANTOM, HPCA, IRIS, Blueprint, and Novershtern) for six major types of immune cells (B cells, CD4+ cells, CD8+ cells, monocytes, neutrophils, NK cells) into cell type-specific vectors, removed duplicated values (genes found in multiple datasets) and calculated proportion of these markers in each of four recurrent MDD-related modules. The results are presented in Table S1.4 and Figure S1.11.

Table S1.4. Number of cell type-specific markers found in recurrent MDD-related modules (N, number of markers in each cell type).

	N	Turquoise	Tan	Black	Lightcyan
B cells	135	22 (16.3%)	2 (1.5%)	1 (0.7%)	1 (0.7%)
CD4+ T cells	160	49 (30.6%)	0 (0%)	0 (0%)	0 (0%)
CD8+ T cells	116	23 (19.8%)	2 (0.9%)	0 (0%)	1 (0.9%)
Monocytes	303	75 (24.8%)	4 (1.3%)	8 (2.6%)	5 (1.7%)
Neutrophils	80	10 (12.5)	1 (1.3%)	0 (0%)	1 (1.3%)
NK cells	100	23 (23%)	1 (1%)	1 (1%)	4 (4%)

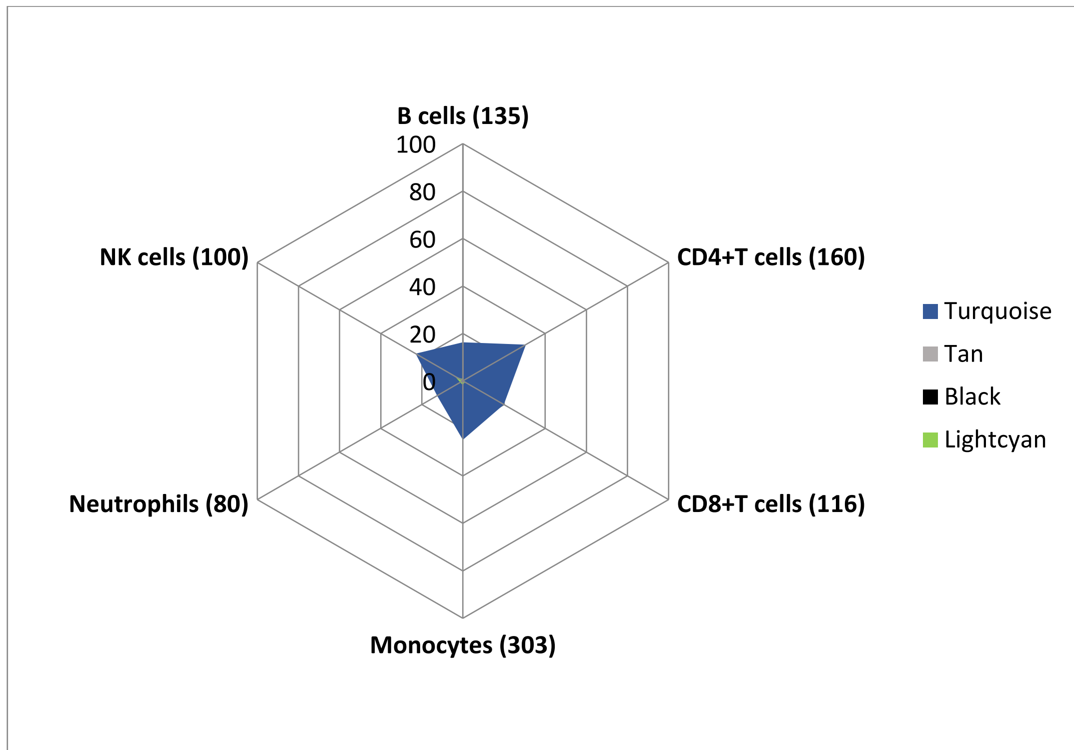


Figure S1.11. Visualisation of relative proportions of genes belonging to recurrent MDD-associated modules found in cell type-specific gene expression signatures.

Given the results presented we suggest that the Turquoise module is likely to be influenced by immune cells-specific gene expression, especially by CD4+ T cells. For the Tan, Black and Lightcyan module we found that negligible number of cell-specific markers (up to 2.6%) appear in these modules, suggesting minimal influence from these cell types.

2.7. Functional analyses

The results of Enrichment (DAVID) and Signalling Pathway Impact Analyses (SPIA) are presented in the **Supplementary Table S7 and Table S8 (online manuscript)**. Graphical representation of each identified pathway can be viewed via KEGG link provided in this table. By clicking on the KEGG link, the pathway graph with the genes that driving dysregulation of this pathway in MAS dataset (highlighted in red) can be accessed.

2.8. Replication (OATS cohort)

RNA purification, Illumina hybridisation, QC checks, pre-processing and filtering in replication cohort (OATS) were performed identically to what was done in the exploratory cohort (sMAS). Moreover, to reinforce high replicability, blood samples collection, RNA extraction and hybridisation were performed using the same protocols, by the same people in the same facilities across two studies.

Assessing comparability of gene expression data

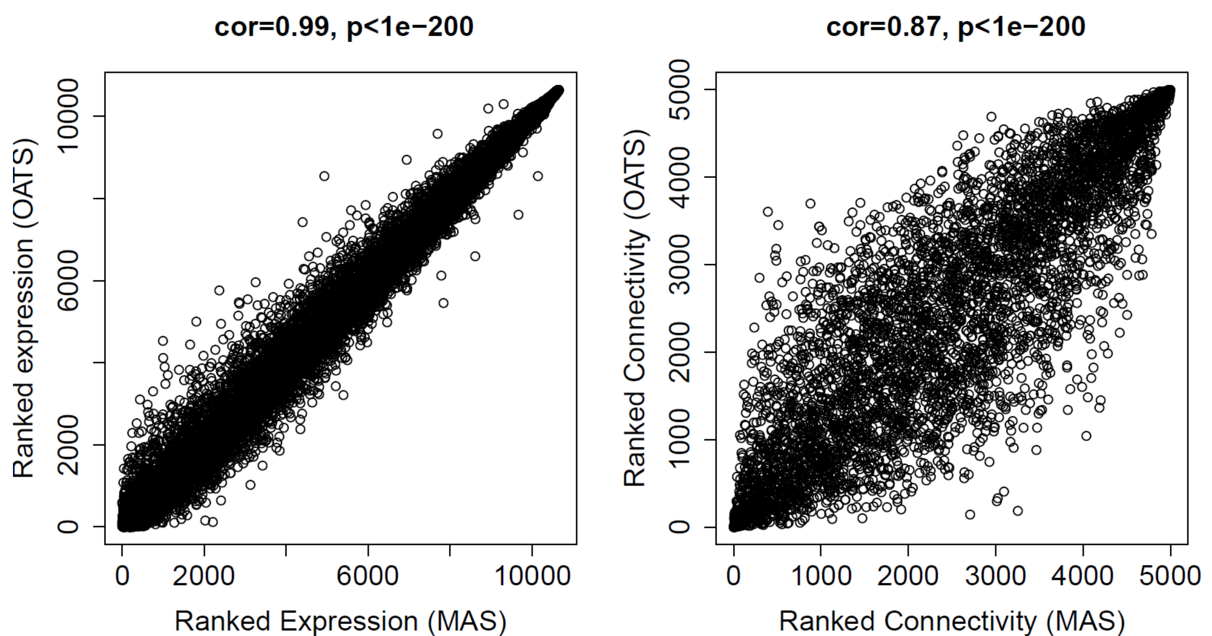


Figure S1.12. General Network properties defined as correlations between ranked expression and ranked connectivity for 10654 genes expressed in both MAS and OATS data sets.

Module preservation

Table S1.5. Module preservation in sMAS and OATS datasets quantified by Z summary score ($5<Z<10$ indicated moderate preservation, $Z>10$ indicates high preservation).

Module	Size (MAS)	Size (OATS)	Z summary
Turquoise	2845	2773	62.02
Tan	315	298	43.14
Black	410	393	42.78
Lightcyan	185	180	31.83

Table S1.6. Replication of MAS findings in OATS. Correlations between the module eigengenes (ME) of Turquoise, Tan, Black, Lightcyan modules and single episode (SE), current (CD), recurrent (RD), and lifetime depression (LD), age, sex and BMI in OATS cohort. Associated p-values are indicated in parenthesis.

Module	Size (% of MAS)	RD r (p-value)	Age r (p-value)	Sex r (p-value)	BMI r (p-value)
Turquoise	2773 (97.5%)	-0.06 (0.4)	-0.13 (0.07)	-0.09 (0.2)	0.4 (0.6)
Tan	298 (94.6%)	0.01 (0.8)	0.02 (0.8)	0.1 (0.2)	0.02 (0.8)
Black	393 (95.9%)	0.03 (0.7)	0.02 (0.8)	0.1 (0.2)	0.02 (0.8)
Lightcyan	180 (97.3%)	0.04 (0.6)	0.03 (0.7)	0.1 (0.1)	-0.05 (0.9)

Table S1.7. Nine common genes identified across three modules in the discovery (MAS) and replication (OATS) cohorts as associated with recurrent MDD in older adults. Module Membership (MM) is computed as correlation of a given module gene with the eigenegene of a module. Gene significance (GS) is a correlation between module eigengene and recurrent depression.

Chr	**SYMBOL** ENTEZ ID	Official full name	Module	MAS		OATS	
				MM	GS r (p-value)	MM	GS r (p-value)
11	*CTSC* 1075	cathepsin C	Turquoise	0.38	-0.14 (0.001)	0.5	-0.17 (0.02)
2	*ORMDL1* 94101	ORMDL sphingolipid biosynthesis regulator 1	Turquoise	0.76	-0.13 (0.003)	0.85	-0.16 (0.03)
13	*NARG1L* 79612	N(alpha)-acetyltransferase 16, NatA auxiliary subunit	Turquoise	0.74	-0.12 (0.008)	0.75	-0.15 (0.046)
3	*B4GALT4* 8702	beta-1,4-galactosyltransferase 4	Turquoise	0.39	-0.11 (0.02)	0.43	-0.16 (0.04)
11	*GTF2H1* 2965	general transcription factor IIH subunit 1	Turquoise	0.34	-0.09 (0.04)	0.32	0.15 (0.046)
10	*AGAP6* 414189	ArfGAP with GTPase domain, ankyrin repeat and PH domain 6	Turquoise	0.46	-0.09 (0.04)	0.44	-0.15 (0.04)
6	*THEMIS* 387357	thymocyte selection associated	Turquoise	0.66	-0.09 (0.04)	0.75	-0.15 (0.04)
3	*IL5RA* 3568	interleukin 5 receptor subunit alpha	Tan	0.31	0.1 (0.03)	0.4	0.2 (0.007)
15	*SNX22* 79856	sorting nexin 22	Lightcyan	0.81	0.1 (0.03)	0.8	0.2 (0.008)

2.9. Significance of the overlap between sMAS and OATS findings

To test for statistical significance of the overlap between recurrent MDD-correlated probes in discovery (sMAS) and replication (OATS) cohorts we calculated the representation factor (RF) and the associated probability.

The groups of probes associated with recurrent MDD were compared in sMAS and OATS found to have x probes in common (Table S).

Representation factor (RF)

The RF is the number of overlapping probes divided by the expected number of overlapping probes drawn from two independent groups.

RF > 1 indicates more overlap than expected of two independent groups;

RF < 1 indicated less overlap than expected of two independent groups.

Probability

Exact hypergeometric test was used to calculate the probability of the overlap between sMAS and OATS recurrent MDD-related probes.

Online tool: http://nemates.org/MA/progs/overlap_stats.cgi

Further details on the calculation can be found at:

http://nemates.org/MA/progs/representation.stats.html

Table S1.8. Statistical significance of the overlap between recurrent MDD-related probes in the discovery (sMAS) and replication (OATS) cohorts.

Module	N of probes correlated with recurrent MDD at p < 0.05		Overlap between sMAS and OATS	Representation factor (RF)	p-value
	sMAS	OATS			
Turquoise	963	38	7	2.1	p < 0.04
Tan	117	1	1	94.2	p < 0.01
Black	99	0	0	NA	NA
Lightcyan	62	1	1	177.7	p < 0.006
Total	**1241**	**40**	**9**	**2.0**	**p < 0.03**

Table S1.9. List of 24 genes belonging to recurrent MDD-associated modules that previously were found as differentially expressed between MDD cases and controls (Leday et al., 2018)

Gene_Symbol	Entrez_ID	Module	GS_recurrent MDD	p_GS_recurrent MDD	q_value	signif_at_FDR_0.2	MM	p_MM
MGST1	4257	Turquoise	-0.02	0.66	0.60	0	0.08	0.083936767
TXN	7295	Turquoise	-0.03	0.57	0.56	0	0.46	3.53E-28
CLEC4A	50856	Turquoise	-0.08	0.08	0.27	0	0.59	3.46E-50
CKLF	51192	Turquoise	-0.03	0.43	0.50	0	0.47	2.00E-30
S100A8	6279	Turquoise	0.01	0.75	0.63	0	0.50	4.53E-34
GABARAP	11345	Turquoise	-0.07	0.09	0.27	0	0.47	1.11E-29
ZBTB4	57659	Turquoise	0.02	0.64	0.59	0	-0.33	4.80E-15
ASXL1	171023	Turquoise	-0.01	0.88	0.67	0	-0.29	1.20E-11
MACF1	23499	Turquoise	0.01	0.86	0.66	0	0.40	1.17E-21
FNBP4	23360	Turquoise	-0.06	0.15	0.33	0	0.61	4.51E-55
TBP	6908	Turquoise	-0.08	0.06	0.24	0	0.34	1.40E-15
PDCD7	10081	Turquoise	0.00	0.93	0.68	0	0.14	0.000972626
SLC7A6	9057	Turquoise	-0.05	0.25	0.40	0	0.35	7.47E-17
RPL22	6146	Turquoise	-0.07	0.13	0.31	0	0.41	6.72E-23
ATP8B2	57198	Turquoise	-0.07	0.11	0.30	0	0.17	0.000125648
NOL8	55035	Turquoise	-0.05	0.25	0.40	0	0.63	6.11E-58
MTSS1	9788	Turquoise	-0.03	0.43	0.50	0	0.54	4.39E-41
TMEM194A	23306	Turquoise	-0.08	0.07	0.25	0	0.64	7.66E-62
ERMP1	79956	Turquoise	-0.02	0.70	0.61	0	0.35	4.09E-16
NFATC3	4775	Tan	0.15	0.00	0.17	1	0.56	1.40E-43
PQLC1	80148	Black	0.05	0.24	0.39	0	0.60	3.20E-53
GDE1	51573	Black	0.06	0.15	0.33	0	0.64	2.15E-60
IMPDH1	3614	Lightcyan	0.09	0.03	0.21	0	0.73	3.88E-89
EPM2AIP1	9852	Lightcyan	-0.04	0.36	0.47	0	-0.56	4.00E-45

Appendix A references:

ALLEN, J. D., CHEN, M. & XIE, Y. 2009. Model-Based Background Correction (MBCB): R Methods and GUI for Illumina Bead-array Data. *Journal of cancer science & therapy,* 1**,** 25-27.

ARAN, D., HU, Z. & BUTTE, A. J. 2017. xCell: digitally portraying the tissue cellular heterogeneity landscape. 18**,** 220.

BOLSTAD, B. M., IRIZARRY, R. A., ASTRAND, M. & SPEED, T. P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics,* 19**,** 185-93.

CUMMINGS, J. L., MEGA, M., GRAY, K., ROSENBERG-THOMPSON, S., CARUSI, D. A. & GORNBEIN, J. 1994. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology,* 44**,** 2308-14.

FISKE, A., WETHERELL, J. L. & GATZ, M. 2009. Depression in Older Adults. *Annual Review of Clinical Psychology,* 5**,** 363-389.

FOLSTEIN, M. F., FOLSTEIN, S. E. & MCHUGH, P. R. 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res,* 12**,** 189-98.

KROENKE, K., SPITZER, R. L. & WILLIAMS, J. B. 2001. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med,* 16**,** 606-13.

LAI, C. K. Y. 2014. The merits and problems of Neuropsychiatric Inventory as an assessment tool in people with dementia and other neurological disorders. *Clinical Interventions in Aging,* 9**,** 1051-1061.

LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics,* 9**,** 559.

LANGFELDER, P. & HORVATH, S. 2012. Fast R Functions for Robust Correlations and Hierarchical Clustering. *Journal of statistical software,* 46**,** i11.

LEDAY, G. G. R., VERTES, P. E., RICHARDSON, S., GREENE, J. R., REGAN, T., KHAN, S., HENDERSON, R., FREEMAN, T. C., PARIANTE, C. M., HARRISON, N. A., PERRY, V. H., DREVETS, W. C., WITTENBERG, G. M. & BULLMORE, E. T. 2018. Replicable and Coupled Changes in Innate and Adaptive Immune Gene Expression in Two Case-Control Studies of Blood Microarrays in Major Depressive Disorder. *Biol Psychiatry,* 83**,** 70-80.

LEEK, J. T., JOHNSON, W. E., PARKER, H. S., JAFFE, A. E. & STOREY, J. D. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics,* 28**,** 882-3.

LEONTJEVAS, R., VAN HOOREN, S. & MULDERS, A. 2009. The Montgomery-Asberg Depression Rating Scale and the Cornell Scale for Depression in Dementia: a validation study with patients exhibiting early-onset dementia. *Am J Geriatr Psychiatry,* 17**,** 56-64.

LIN, S. M., DU, P., HUBER, W. & KIBBE, W. A. 2008. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Research,* 36**,** e11.

SACHDEV, P. S., BRODATY, H., REPPERMUND, S., KOCHAN, N. A., TROLLOR, J. N., DRAPER, B., SLAVIN, M. J., CRAWFORD, J., KANG, K., BROE, G. A., MATHER, K. A. & LUX, O. 2010. The Sydney Memory and Ageing Study (MAS): methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70-90 years. *Int Psychogeriatr,* 22**,** 1248-64.

SACHDEV, P. S., LAMMEL, A., TROLLOR, J. N., LEE, T., WRIGHT, M. J., AMES, D., WEN, W., MARTIN, N. G., BRODATY, H. & SCHOFIELD, P. R. 2012. A

Comprehensive Neuropsychiatric Study of Elderly Twins: The Older Australian Twins Study. *Twin Research and Human Genetics,* 12**,** 573-582.
XIE, Y., WANG, X. & STORY, M. 2009. Statistical methods of background correction for Illumina BeadArray data. *Bioinformatics,* 25**,** 751-7.
YESAVAGE, J. A., BRINK, T. L., ROSE, T. L., LUM, O., HUANG, V., ADEY, M. & LEIRER, V. O. 1982. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res,* 17**,** 37-49.

Appendix B

Supplementary materials for the manuscript "Downregulated transferrin receptor in the blood predicts recurrent MDD in the elderly cohort: a fuzzy forests approach"

Our motivation for choosing fuzzy forests algorithm

Fuzzy forests is an algorithm that combines two frameworks: co-expression based hierarchical clustering implemented in WGCNA and ensemble classifier Random forests. First, data is clustered into relatively uncorrelated modules of highly correlated features within the modules. Then, the most important features are selected using random forests recursive feature elimination in each module separately. The last iteration of random forests includes all survived features across all modules to build a final predictive model. The motivation for choosing this algorithm is twofold. Firstly, transcriptome data is highly multidimensional with a large proportion of correlated features. Clustering prior fitting random forests into clusters help to alleviate multicollinearity problem to which random forests is known to be sensitive to. Secondly, fuzzy logic implemented in the algorithm reduces dimensionality to alleviate overfitting due to p>>n problem. Given that the dataset was heavily unbalanced, we balanced our testing data using random undersampling of the majority class as well as k-nearest oversampling implemented in SMOTE algorithm.

Overview of WGCNA

WGCNA was developed to detect the correlational structure in biological networks, assuming that genes with high correlations are likely to be involved in the same biological processes. The R package WGCNA, a well-documented implementation of the WGCNA framework, has shown a great success in many biological applications, including our previous work (Langfelder and Horvath, 2008). Briefly, to construct a network, we first define a similarity function. This similarity function is based on Pearson's' correlation calculated for each possible pair of genes in transcriptome. To define the connection

strength measure between the genes in the network, these correlations are weighted by taking the absolute value and rising to the power β. Network connectivity is defined as a sum of connection strengths for each gene, describing how strongly each gene is connected to the other genes in the network. Next, we calculate the topological overlap for each pair of genes and identify groups of genes with high topological overlap, where both genes in a pair are strongly connected to the same group of genes. Hierarchical clustering is used to identify clusters, or modules, of interconnected genes (Zhang and Horvath, 2005).

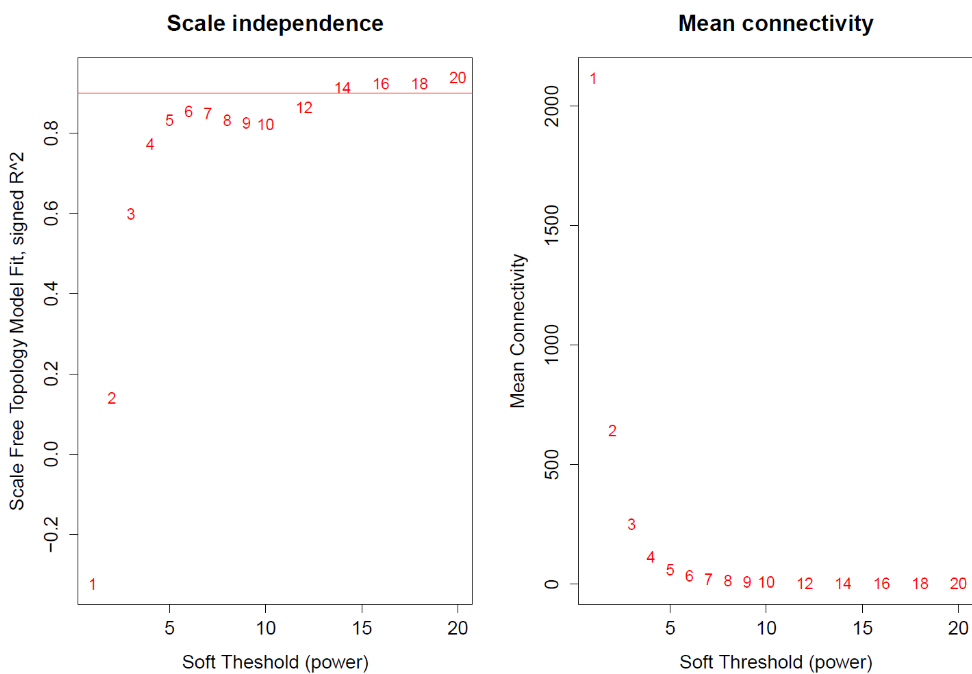


Figure S2.1. Scale independence and mean connectivity**.** Analysis of network topology for various soft-thresholding powers. The left panel shows the scale-free fit index (y-axis) as a function of the soft-thresholding power (x-axis). The right panel displays the mean connectivity (degree, y-axis) as a function of the soft-thresholding power (x-axis).

Random forests (RF) and Variable Importance Measures (VIM)

RF is an ensemble method that works by combining the predictions of an ensemble of classification (or regression) decision trees introduced by Breiman (2001). The principle of RF is to combine many binary decision trees built using several bootstrap samples obtained from the training sample. Each tree is grown on a separate bootstrap sample of the training data. About one third of observations that are not selected in a particular

bootstrap sample are out of bag (OOB) for each tree. By averaging prediction error over multiple trees and many bootstrap samples the estimated prediction error was obtained to assess model performance.

If the goal is to build a predictive model using transcriptome data that can be utilized for diagnostic purposes, selection of the most relevant features out of multidimensional feature space is necessary. This requires an understanding of how each individual feature affects the model, which can be evaluated by Variable Importance Measures, VIMs. We estimated VIMs by calculating the effect of random permutations of the values of an individual feature (standardized level of gene expression) on predictive accuracy of the target outcome on test data, i.e. VIMs estimate the average decline in predictive performance for each individual feature across multiple trees, therefore providing reliable measure of variable importance for the prediction of outcome. RF VIMs were calculated for the final model. Given that random forests are non-linear and non-parametric, VIMs, defined above, thus, naturally take into account non-linear interactions between the genes. Score of importance of a given variable was calculated as *Mean Decrease Accuracy* and *Mean Decrease Gini* coefficients.

The effect of the number of pre-selected features for the final prediction model

The number of features included in the final model is a manually specified the *number_selected* parameter. We observed that while the ranking order of the features were relatively stable for different numbers of features selected the performance of the model was greatly affected by the number_selected parameter. As an example, Figure S2.2 shows variable importance for the number_selected = 10 as measured by Mean Decrease Accuracy and Mean Decrease Gini indices. As can be seen, the *TFRC* is the top predictor of the target outcome, which alone predicted the outcome with sensitivity of 63%; however, we were unable to achieve clinically meaningful sensitivity using multiple features.

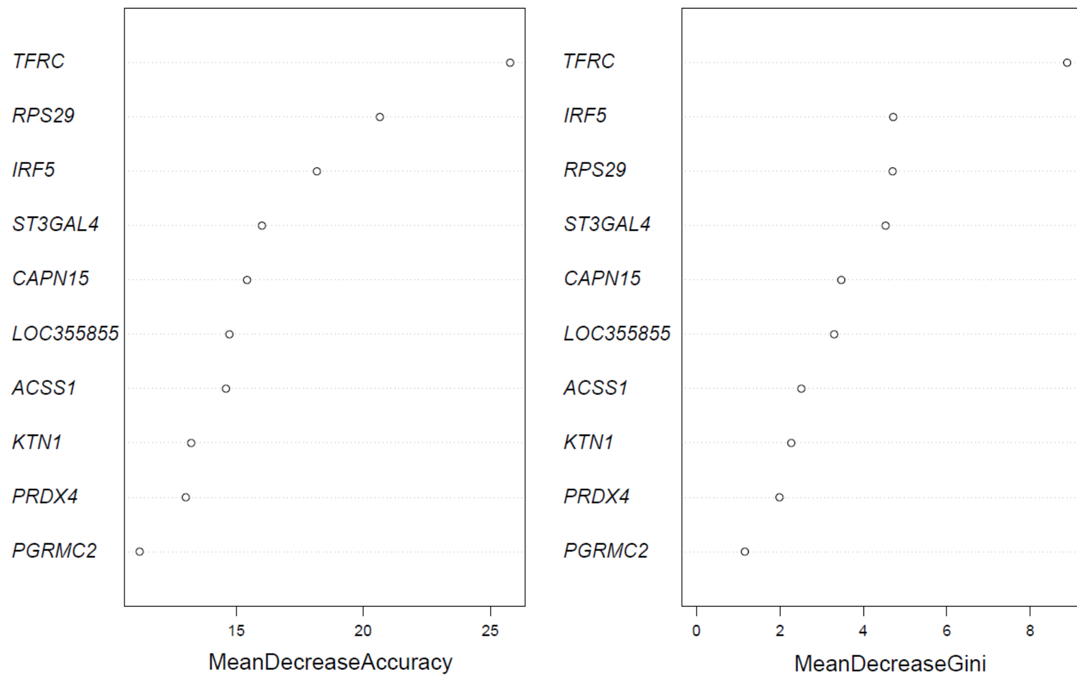


Figure S2.2. Variable importance plot for 10 most predictive features. The plot shows each variable on the y-axis, and their importance on the x-axis.

Fuzzy forests algorithm

FF, being an extension of random forests, is powerful algorithm that was proposed to deal with correlated, high dimensional data (Conn et al., 2015, Conn et al., 2016). FF works using piecewise approach. First, the network structure of the data and partition of features into distinct modules such that the correlation within each module is high and the correlation between modules is low is estimated using WGCNA (Zhang and Horvath, 2005). FF reduces dimensionality of the data in two steps: a screening step and a selection step. The screening step uses recursive feature elimination random forests (RFE-RF) (Díaz-Uriarte and Alvarez de Andrés, 2006) independently on each module to screen out unimportant features assigned to the module by WGCNA. Given that we apply this algorithm to biological data, it is reasonable to assume that while modules are relatively independent from each other, there is potential for interaction between the modules. The selection step is the final iteration of RFE-RF, which was applied to all features from all modules that have been selected at the screening step. Thus, the potential correlation bias between the features is re-introduced, allowing, therefore, for interactions between features that were found to be important within individual modules.

Lastly, based on a specified number of a ranked variable importance list a final group of selected features is used to train a predictive model.

To optimize the model performance, FF allows for tuning several parameters: *drop_fraction* (number of features to drop at each iteration), *keep_fraction* (number of features to retain at each iteration), *number_selected* (number of important features as the output of the final random forest), *ntree_factor, min_ntree* (minimum number of trees grown in each random forest), and *final_ntree* (number of trees grown in the final random forest containing all selected features). Since the number of features varies across random forests, for each random forest, parameters *mtry* (number of variables randomly sampled at each split) and *ntree* (number of trees to grow) were specified as a function of the current number of features:

$$mtry = ceiling((p/3)*mtry_factor,$$

where *p* is a number of features in a given module and *mtry_factor* = 5;

The parameter *ntree* for each random forest is *ntree_factor* times the number of features:

$$ntree = max(min_ntree, ntree_factor*p),$$

where *min_ntree = 5* and *ntree_factor = 5.*

We set *drop_fraction* = 0.25 and *keep_fraction* = 0.05; *number_selected = 1, 3, 5, 10, 20, 50, or 100* and *min_ntree* = 100, 500, 1000, 5000, 10000.

The *number_selected* parameter, which is the number of features to be determined as the top predictors at the RF iteration, was set to 1, 3, 5, 10, 20, 50, 100.

The algorithm was implemented using R package '*fuzzyforest'*.

References

BREIMAN, L. 2001. Random Forests. *Machine Learning,* 45**,** 5-32.

CONN, D., NGUN, T., LI, G. & RAMIREZ, C. 2015. Fuzzy Forests: Extending Random Forests for Correlated, High-Dimensional Data. *UCLA: Biostatistics*.

CONN, D., NGUN, T., LI, G. & RAMIREZ, C. M. 2016. Fuzzy forests: a new WGCNA based random forest algorithm for correlated, high-dimensional data. *Journal of Statistical Software*.

DÍAZ-URIARTE, R. & ALVAREZ DE ANDRÉS, S. 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics,* 7**,** 3-3.

GAMMELLA, E., BURATTI, P., CAIRO, G. & RECALCATI, S. 2017. The transferrin receptor: the cellular iron gate. *Metallomics,* 9**,** 1367-1375.

LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics,* 9**,** 559.

LEE, P. L. & BEUTLER, E. 2009. Regulation of hepcidin and iron-overload disease. *Annu Rev Pathol,* 4**,** 489-515.

MAKKER, J., HANIF, A., BAJANTRI, B. & CHILIMURI, S. 2015. Dysmetabolic Hyperferritinemia: All Iron Overload Is Not Hemochromatosis. *Case Reports in Gastroenterology,* 9**,** 7-14.

PORTO, G., BRISSOT, P., SWINKELS, D. W., ZOLLER, H., KAMARAINEN, O., PATTON, S., ALONSO, I., MORRIS, M. & KEENEY, S. 2015. EMQN best practice guidelines for the molecular genetic diagnosis of hereditary hemochromatosis (HH). *European Journal Of Human Genetics,* 24**,** 479.

ZHANG, B. & HORVATH, S. 2005. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol,* 4**,** Article17.

Appendix C

Table S3.1. The top 100 pathogen- and immune-related genes in association with recurrent MDD

ILMN_probe	Gene symbol	Entrez_ID	logFC	AveExpr	P-value	FDR 5%
ILMN_1740015	CD14	929	-0.69448	9.90586	9.51E-07	0.000916
ILMN_2396444	CD14	929	-0.6436	12.15393	3.19E-06	0.001535
ILMN_2070300	LSM2	57819	0.263971	8.561128	0.000563	0.157371
ILMN_2152131	ACTB	60	-0.29085	13.05806	0.000654	0.157371
ILMN_1787705	ATP6V1B2	526	-0.33626	11.03821	0.000882	0.169787
ILMN_1658121	CFP	5199	-0.30173	7.950343	0.00109	0.174956
ILMN_1767365	PAK1	5058	-0.29518	8.469129	0.001564	0.215213
ILMN_1700428	HLA-DOB	3112	0.310135	7.338196	0.002755	0.261859
ILMN_1799134	KLRD1	3824	0.446587	8.018866	0.002999	0.261859
ILMN_1729987	SRC	6714	-0.15924	7.190677	0.003263	0.261859
ILMN_1782538	VIM	7431	-0.34665	11.5966	0.003358	0.261859
ILMN_1728799	FBP1	2203	-0.26878	8.140827	0.004249	0.261859
ILMN_1690546	PPP3CC	5533	0.178654	7.416294	0.004786	0.261859
ILMN_1778977	TYROBP	7305	-0.30005	12.60216	0.004816	0.261859
ILMN_1654396	ITGB2	3689	-0.29333	12.80852	0.004918	0.261859
ILMN_1789074	HSPA1A	3303	-0.45897	9.398591	0.005326	0.261859
ILMN_2386790	KLRC3	3823	0.17302	6.915689	0.005448	0.261859
ILMN_1663160	ZNF337	26152	0.14795	7.634455	0.005534	0.261859
ILMN_1799725	DOCK2	1794	-0.24106	9.758014	0.005846	0.261859
ILMN_1740493	TRAF5	7188	0.17357	7.39066	0.006153	0.261859
ILMN_2038777	ACTB	60	-0.30301	12.83257	0.006362	0.261859
ILMN_2043816	ARPC5L	81873	0.147105	7.50646	0.006929	0.261859
ILMN_2175912	ITGB2	3689	-0.28925	12.81918	0.006991	0.261859
ILMN_1783709	RRAGA	10670	0.137455	8.49008	0.007355	0.261859
ILMN_1738523	MYD88	4615	-0.19647	8.357946	0.0074	0.261859
ILMN_1777220	VCP	7415	-0.14859	7.665732	0.007598	0.261859
ILMN_1736577	ZNF688	146542	0.078748	7.044594	0.007691	0.261859
ILMN_1743646	VASP	7408	-0.31046	8.125938	0.008101	0.261859
ILMN_1785179	UBE2G2	7327	0.169525	7.738553	0.008508	0.261859
ILMN_2156172	HK2	3099	-0.09421	7.041178	0.008652	0.261859
ILMN_1677483	EXOSC1	51013	0.091585	7.206799	0.00895	0.261859
ILMN_1797988	KLRD1	3824	0.385149	7.950138	0.008956	0.261859
ILMN_1727284	CD4	920	-0.14903	7.437383	0.008973	0.261859
ILMN_2058251	VIM	7431	-0.28734	10.76031	0.009289	0.263093
ILMN_2310589	DIABLO	56616	0.114487	8.429549	0.009713	0.267237
ILMN_1661554	DIAPH1	1729	-0.21661	7.877845	0.010864	0.280583
ILMN_1670302	HK3	3101	-0.25128	8.298209	0.010915	0.280583
ILMN_1710756	ENO1	2023	-0.23966	10.80819	0.011072	0.280583
ILMN_1666269	CTSZ	1522	-0.21966	7.771136	0.013034	0.313932
ILMN_1801105	PRKCD	5580	-0.25181	9.674262	0.014312	0.313932
ILMN_1782704	CD19	930	0.364886	7.412291	0.014432	0.313932
ILMN_1778360	PYGB	5834	-0.17446	7.535417	0.014624	0.313932
ILMN_2399392	SIL1	64374	-0.07433	6.901988	0.014731	0.313932

ILMN_1814465	UBE2G1	7326	0.136778	7.388585	0.014741	0.313932
ILMN_1738326	EIF4E2	9470	-0.12767	8.235022	0.015087	0.313932
ILMN_2397954	PARP3	10039	0.117501	7.403345	0.01512	0.313932
ILMN_1696187	PYGL	5836	-0.32474	9.275016	0.015322	0.313932
ILMN_1779177	U2AF1L4	199746	0.085796	7.138459	0.016295	0.323588
ILMN_1723486	HK2	3099	-0.15133	7.28403	0.016465	0.323588
ILMN_1657483	SEC23B	10483	-0.11031	7.783979	0.017062	0.328612
ILMN_3300358	ZNF84	7637	0.082691	6.942481	0.017796	0.329452
ILMN_2312149	ZNF160	90338	0.074512	6.976155	0.018338	0.329452
ILMN_2360784	RRBP1	6238	-0.18561	7.968479	0.018348	0.329452
ILMN_1704404	PSMD13	5719	-0.06521	6.989597	0.018474	0.329452
ILMN_2209748	DERL1	79139	0.177152	7.729297	0.019163	0.331618
ILMN_1698419	NCOR2	9612	-0.15518	7.509525	0.019544	0.331618
ILMN_2126706	LMNB1	4001	-0.09505	7.016712	0.019628	0.331618
ILMN_1675674	UBE4B	10277	-0.13078	7.85877	0.020035	0.332654
ILMN_1812403	BCAP31	10134	-0.16238	9.093416	0.020957	0.337308
ILMN_1766275	PIK3CD	5293	-0.24303	8.722367	0.021585	0.337308
ILMN_1758105	ZNF791	163049	0.107228	7.383606	0.021666	0.337308
ILMN_1729915	PILRA	29992	-0.23221	8.236452	0.021774	0.337308
ILMN_1762003	SEC62	7095	0.108246	7.140513	0.022067	0.337308
ILMN_2367020	SEC61G	23480	0.200263	8.986863	0.023566	0.35459
ILMN_2241953	PILRA	29992	-0.22745	8.102257	0.025831	0.38269
ILMN_1807277	IFI30	10437	-0.24818	9.018814	0.027117	0.39566
ILMN_1795822	DIS3L	115752	0.136976	7.701837	0.028234	0.401373
ILMN_1812915	TNFRSF10B	8795	-0.34032	8.652337	0.02867	0.401373
ILMN_1675788	ZNF175	7728	0.062265	6.902299	0.029211	0.401373
ILMN_1777049	ZNF160	90338	0.157271	7.496851	0.029727	0.401373
ILMN_1674160	BIN1	274	0.205495	8.66141	0.030745	0.401373
ILMN_2334242	CREB1	1385	0.287781	9.812279	0.031245	0.401373
ILMN_1680693	ZNF419	79744	0.080125	7.474288	0.031286	0.401373
ILMN_1662451	FCER2	2208	0.217731	7.163268	0.031599	0.401373
ILMN_1674038	CTSD	1509	-0.25868	8.23181	0.031892	0.401373
ILMN_2321416	DIAPH1	1729	-0.2275	8.913148	0.032941	0.401373
ILMN_2083469	IRS2	8660	-0.22138	7.577877	0.033198	0.401373
ILMN_2313821	AIFM1	9131	-0.06945	7.348443	0.033505	0.401373
ILMN_1744980	ZCCHC7	84186	0.153612	7.342094	0.033782	0.401373
ILMN_2309245	BIN1	274	0.190631	8.510109	0.033928	0.401373
ILMN_1653711	FZD2	2535	-0.08769	6.955879	0.033939	0.401373
ILMN_1762825	CANX	821	-0.12807	7.835239	0.034293	0.401373
ILMN_1678962	DFFB	1677	0.066668	7.052557	0.034694	0.401373
ILMN_1660533	RPN1	6184	-0.15075	9.445104	0.035011	0.401373
ILMN_1772113	U2AF1	7307	-0.08878	7.575507	0.035827	0.403685
ILMN_1768194	BIRC2	329	0.073412	6.985571	0.036593	0.403685
ILMN_1727402	HCLS1	3059	-0.22965	11.40675	0.036762	0.403685
ILMN_1781290	RHOA	387	-0.19906	11.16215	0.036889	0.403685
ILMN_1787026	SEC61G	23480	0.192212	9.250013	0.03857	0.414548
ILMN_1678919	YOD1	55432	0.281611	7.356455	0.038743	0.414548
ILMN_1801928	YWHAZ	7534	-0.17459	11.06982	0.040615	0.428025
ILMN_1781996	NUDT16	131870	-0.06442	6.989968	0.040891	0.428025
ILMN_1779735	C7ORF59	389541	-0.17879	9.873468	0.041546	0.430202
ILMN_1685365	ZNF773	374928	0.075789	7.004205	0.042572	0.433188
ILMN_1733324	ITGB3	3690	-0.1723	7.03312	0.042734	0.433188
ILMN_2131861	SOCS2	8835	0.130674	7.24166	0.043526	0.4362

ILMN_1665964	GAB2	9846	-0.22659	8.423128	0.043937	0.4362
ILMN_1699265	TNFRSF10B	8795	-0.12544	7.987933	0.045314	0.445277
ILMN_1742521	GRB2	2885	-0.1251	7.907487	0.045855	0.446003
ILMN_1739792	RHOG	391	-0.24455	11.40908	0.046314	0.446003

Bibliography

ABRAHAM, G., KOWALCZYK, A., ZOBEL, J. & INOUYE, M. 2013. Performance and Robustness of Penalized and Unpenalized Methods for Genetic Prediction of Complex Human Disease. *Genetic Epidemiology,* 37**,** 184-195.

AKULA, N., WENDLAND, J. R., CHOI, K. H. & MCMAHON, F. J. 2015. An Integrative Genomic Study Implicates the Postsynaptic Density in the Pathogenesis of Bipolar Disorder. *Neuropsychopharmacology,* 41**,** 886.

ALLEN, J. D., CHEN, M. & XIE, Y. 2009. Model-Based Background Correction (MBCB): R Methods and GUI for Illumina Bead-array Data. *Journal of cancer science & therapy,* 1**,** 25-27.

ALTMAN, D. G. 2006. *Practical Statistics for Medical Research*, Chapman \& Hall/CRC.

AMAR, D., SAFER, H. & SHAMIR, R. 2013. Dissection of Regulatory Networks that Are Altered in Disease via Differential Co-expression. *PLOS Computational Biology,* 9**,** e1002955.

AMBROSIO, G., KAUFMANN, F. N., MANOSSO, L., PLATT, N., GHISLENI, G., RODRIGUES, A. L. S., RIEGER, D. K. & KASTER, M. P. 2018. Depression and peripheral inflammatory profile of patients with obesity. *Psychoneuroendocrinology,* 91**,** 132-141.

ANISMAN, H., MERALI, Z. & POULTER, M. O. 2012. *Gamma-Aminobutyric Acid Involvement in Depressive Illness Interactions with Corticotropin-Releasing Hormone and Serotonin*

The Neurobiological Basis of Suicide, Boca Raton FL, Llc.

ANTONENKOV, V. D. & HILTUNEN, J. K. 2012. Transfer of metabolites across the peroxisomal membrane. *Biochim Biophys Acta,* 1822**,** 1374-86.

ANTYPA, N., SERRETTI, A. & RUJESCU, D. 2013. Serotonergic genes and suicide: a systematic review. *Eur Neuropsychopharmacol,* 23**,** 1125-42.

ARAN, D., HU, Z. & BUTTE, A. J. 2017. xCell: digitally portraying the tissue cellular heterogeneity landscape. 18**,** 220.

ASTON, C., JIANG, L. & SOKOLOV, B. P. 2005. Transcriptional profiling reveals evidence for signaling and oligodendroglial abnormalities in the temporal cortex from patients with major depressive disorder. *Mol Psychiatry,* 10**,** 309-22.

BAKUNINA, N., PARIANTE, C. M. & ZUNSZAIN, P. A. 2015. Immune mechanisms linked to depression via oxidative stress and neuroprogression. *Immunology*.

BAUNE, B. T. 2018. Chapter 35 - Future Perspectives on Immune-Related Treatments. *In:* BAUNE, B. T. (ed.) *Inflammation and Immunity in Depression.* Academic Press.

BELZEAUX, R., BERGON, A., JEANJEAN, V., LORIOD, B., FORMISANO-TREZINY, C., VERRIER, L., LOUNDOU, A., BAUMSTARCK-BARRAU, K., BOYER, L., GALL, V., GABERT, J., NGUYEN, C., AZORIN, J. M., NAUDIN, J. & IBRAHIM, E. C. 2012. Responder and nonresponder patients exhibit different peripheral transcriptional signatures during major depressive episode. *Transl Psychiatry,* 2**,** e185.

BELZEAUX, R., FORMISANO-TRÉZINY, C., LOUNDOU, A., BOYER, L., GABERT, J., SAMUELIAN, J. C., FÉRON, F., NAUDIN, J. & IBRAHIM, E. C. 2010. Clinical variations modulate patterns of gene expression and define blood biomarkers in major depression. *Journal of Psychiatric Research,* 44**,** 1205-1213.

BELZEAUX, R., LIN, C.-W., DING, Y., BERGON, A., IBRAHIM, E. C., TURECKI, G., TSENG, G. & SIBILLE, E. 2016. Predisposition to treatment response in major depressive episode: A peripheral blood gene coexpression network analysis. *Journal of Psychiatric Research,* 81**,** 119-126.

BERNARD, R., KERMAN, I. A., THOMPSON, R. C., JONES, E. G., BUNNEY, W. E., BARCHAS, J. D., SCHATZBERG, A. F., MYERS, R. M., AKIL, H. & WATSON, S. J. 2011. Altered expression of glutamate signaling, growth factor, and glia genes in the locus coeruleus of patients with major depression. *Mol Psychiatry,* 16**,** 634-46.

BERSANELLI, M., MOSCA, E., REMONDINI, D., GIAMPIERI, E., SALA, C., CASTELLANI, G. & MILANESI, L. 2016. Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinformatics,* 17**,** S15.

BOLSTAD, B. M., IRIZARRY, R. A., ASTRAND, M. & SPEED, T. P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics,* 19**,** 185-93.

BONDY, B. 2007. Common genetic factors for depression and cardiovascular disease. *Dialogues Clin Neurosci,* 9**,** 19-28.

BRANDAO, M., OLIVEIRA, J. C., BRAVO, F., REIS, J., GARRIDO, I. & PORTO, G. 2005. The soluble transferrin receptor as a marker of iron homeostasis in normal subjects and in HFE-related hemochromatosis. *Haematologica,* 90**,** 31-7.

BREEN, M. S., WHITE, C. H., SHEKHTMAN, T., LIN, K., LOONEY, D., WOELK, C. H. & KELSOE, J. R. 2016. Lithium-responsive genes and gene networks in bipolar disorder patient-derived lymphoblastoid cell lines. *The Pharmacogenomics Journal,* 16**,** 446.

BREIMAN, L. 2001. Random Forests. *Machine Learning,* 45**,** 5-32.

CAI, N., CHANG, S., LI, Y., LI, Q., HU, J., LIANG, J., SONG, L., KRETZSCHMAR, W., GAN, X., NICOD, J., RIVERA, M., DENG, H., DU, B., LI, K., SANG, W., GAO, J., GAO, S., HA, B., HO, H. Y., HU, C., HU, J., HU, Z., HUANG, G., JIANG, G., JIANG, T., JIN, W., LI, G., LI, K., LI, Y., LI, Y., LI, Y., LIN, Y. T., LIU, L., LIU, T., LIU, Y., LIU, Y., LU, Y., LV, L., MENG, H., QIAN, P., SANG, H., SHEN, J., SHI, J., SUN, J., TAO, M., WANG, G., WANG, G., WANG, J., WANG, L., WANG, X., WANG, X., YANG, H., YANG, L., YIN, Y., ZHANG, J., ZHANG, K., SUN, N., ZHANG, W., ZHANG, X., ZHANG, Z., ZHONG, H., BREEN, G., WANG, J., MARCHINI, J., CHEN, Y., XU, Q., XU, X., MOTT, R., HUANG, G. J., KENDLER, K. & FLINT, J. 2015. Molecular signatures of major depression. *Curr Biol,* 25**,** 1146-56.

CARDINALI, D. P., SRINIVASAN, V., BRZEZINSKI, A. & BROWN, G. M. 2013. Melatonin and its analogs in insomnia and depression. *Journal of Pineal Research,* 52**,** 365-375.

CARTER, C. J. 2013. Susceptibility genes are enriched in those of the herpes simplex virus 1/host interactome in psychiatric and neurological disorders. *Pathogens and Disease,* 69**,** 240-261.

CARVALHO, L. A., BERGINK, V., SUMASKI, L., WIJKHUIJS, J., HOOGENDIJK, W. J., BIRKENHAGER, T. K. & DREXHAGE, H. A. 2014. Inflammatory activation is associated with a reduced glucocorticoid receptor alpha/beta expression ratio in monocytes of inpatients with melancholic major depressive disorder. *Transl Psychiatry,* 4**,** e344.

CASSANO, P. & FAVA, M. 2002. Depression and public health: an overview. *J Psychosom Res,* 53**,** 849-57.

CATTANE, N., MINELLI, A., MILANESI, E., MAJ, C., BIGNOTTI, S., BORTOLOMASI, M., CHIAVETTO, L. B. & GENNARELLI, M. 2015. Altered Gene Expression in Schizophrenia: Findings from Transcriptional Signatures in Fibroblasts and Blood. *PLoS ONE,* 10**,** e0116686.

CATTANEO, A., MACCHI, F., PLAZZOTTA, G., VERONICA, B., BOCCHIO-CHIAVETTO, L., RIVA, M. A. & PARIANTE, C. M. 2015. Inflammation and neuronal plasticity: a link between childhood trauma and depression pathogenesis. *Frontiers in Cellular Neuroscience,* 9.

CHA, K., HWANG, T., OH, K. & YI, G.-S. 2015. Discovering transnosological molecular basis of human brain diseases using biclustering analysis of integrated gene expression data. *BMC Medical Informatics and Decision Making,* 15**,** S7-S7.

CHARI, R., THU, K. L., WILSON, I. M., LOCKWOOD, W. W., LONERGAN, K. M., COE, B. P., MALLOFF, C. A., GAZDAR, A. F., LAM, S., GARNIS, C., MACAULAY, C. E., ALVAREZ, C. E. & LAM, W. L. 2010. Integrating the multiple dimensions of genomic and epigenomic landscapes of cancer. *Cancer metastasis reviews,* 29**,** 73-93.

CHARLESWORTH, J. C., CURRAN, J. E., JOHNSON, M. P., GORING, H. H., DYER, T. D., DIEGO, V. P., KENT, J. W., JR., MAHANEY, M. C., ALMASY, L., MACCLUER, J. W., MOSES, E. K. & BLANGERO, J. 2010. Transcriptomic epidemiology of smoking: the effect of smoking on gene expression in lymphocytes. *BMC Med Genomics,* 3**,** 29.

CHAWLA, N. V., BOWYER, K. W., HALL, L. O. & KEGELMEYER, W. P. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research,* 16**,** 321-357.

CHEN, C., CHENG, L., GRENNAN, K., PIBIRI, F., ZHANG, C., BADNER, J. A., GERSHON, E. S. & LIU, C. 2013. Two Gene Co-expression Modules Differentiate Psychotics and Controls. *Molecular psychiatry,* 18**,** 1308-1314.

CHISHOLM, D., SWEENY, K., SHEEHAN, P., RASMUSSEN, B., SMIT, F., CUIJPERS, P. & SAXENA, S. 2016. Scaling-up treatment of depression and anxiety: a global return on investment analysis. *The Lancet Psychiatry,* 3**,** 415-424.

CHOUDARY, P. V., MOLNAR, M., EVANS, S. J., TOMITA, H., LI, J. Z., VAWTER, M. P., MYERS, R. M., BUNNEY, W. E., JR., AKIL, H., WATSON, S. J. & JONES, E. G. 2005. Altered cortical glutamatergic and GABAergic signal transmission with glial involvement in depression. *Proc Natl Acad Sci U S A,* 102**,** 15653-8.

CHU, T. T., LIU, Y. & KEMETHER, E. 2009. Thalamic transcriptome screening in three psychiatric states. *Journal of Human Genetics,* 54**,** 665-675.

CIOBANU, L. G. & BAUNE, B. T. 2018. Chapter 11 - Gene Expression of Inflammation Markers in Depression. *In:* BAUNE, B. T. (ed.) *Inflammation and Immunity in Depression.* Academic Press.

CIOBANU, L. G., FERRARI, A. J., ERSKINE, H. E., SANTOMAURO, D. F., CHARLSON, F. J., LEUNG, J., AMARE, A. T., OLAGUNJU, A. T., WHITEFORD, H. A. & BAUNE, B. T. 2018a. The prevalence and burden of mental and substance use disorders in Australia: Findings from the Global Burden of Disease Study 2015. *Australian & New Zealand Journal of Psychiatry,* 52**,** 483-490.

CIOBANU, L. G., SACHDEV, P. S., TROLLOR, J. N., REPPERMUND, S., THALAMUTHU, A., MATHER, K. A., COHEN-WOODS, S. & BAUNE, B. T. 2016. Differential gene expression in brain and peripheral tissues in depression across the life span: A review of replicated findings. *Neuroscience & Biobehavioral Reviews,* 71**,** 281-293.

CIOBANU, L. G., SACHDEV, P. S., TROLLOR, J. N., REPPERMUND, S., THALAMUTHU, A., MATHER, K. A., COHEN-WOODS, S., STACEY, D., TOBEN, C., SCHUBERT, K. O. & BAUNE, B. T. 2018b. Co-expression network analysis of peripheral blood transcriptome identifies dysregulated protein processing in endoplasmic reticulum and immune response in recurrent MDD in older adults. *Journal of Psychiatric Research.*

CIPRIANI, A., FURUKAWA, T. A., SALANTI, G., CHAIMANI, A., ATKINSON, L. Z., OGAWA, Y., LEUCHT, S., RUHE, H. G., TURNER, E. H., HIGGINS, J. P. T., EGGER, M., TAKESHIMA, N., HAYASAKA, Y., IMAI, H., SHINOHARA, K., TAJIKA, A., IOANNIDIS, J. P. A. & GEDDES, J. R. 2018. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with

major depressive disorder: a systematic review and network meta-analysis. *Lancet,* 391**,** 1357-1366.

CIUCULETE, D. M., BOSTROM, A. E., VOISIN, S., PHILIPPS, H., TITOVA, O. E., BANDSTEIN, M., NIKONTOVIC, L., WILLIAMS, M. J., MWINYI, J. & SCHIOTH, H. B. 2017. A methylome-wide mQTL analysis reveals associations of methylation sites with GAD1 and HDAC3 SNPs and a general psychiatric risk score. *Transl Psychiatry,* 7**,** e1002.

CLAPP, M., AURORA, N., HERRERA, L., BHATIA, M., WILEN, E. & WAKEFIELD, S. 2017. Gut microbiota's effect on mental health: The gut-brain axis. *Clinics and Practice,* 7**,** 987.

CONN, D., NGUN, T., LI, G. & RAMIREZ, C. 2015. Fuzzy Forests: Extending Random Forests for Correlated, High-Dimensional Data. *UCLA: Biostatistics*.

CONN, D., NGUN, T., LI, G. & RAMIREZ, C. M. 2016. Fuzzy forests: a new WGCNA based random forest algorithm for correlated, high-dimensional data. *Journal of Statistical Software*.

CONVERGE 2015. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature,* 523**,** 588-591.

COURTET, P. & OLIE, E. 2012. Circadian dimension and severity of depression. *Eur Neuropsychopharmacol,* 22 Suppl 3**,** S476-81.

CRAMER, A. O. J., VAN BORKULO, C. D., GILTAY, E. J., VAN DER MAAS, H. L. J., KENDLER, K. S., SCHEFFER, M. & BORSBOOM, D. 2016. Major Depression as a Complex Dynamic System. *PLoS ONE,* 11**,** e0167490.

CUMMINGS, J. L., MEGA, M., GRAY, K., ROSENBERG-THOMPSON, S., CARUSI, D. A. & GORNBEIN, J. 1994a. The Neuropsychiatric Inventory. *Comprehensive assessment of psychopathology in dementia,* 44**,** 2308-2308.

CUMMINGS, J. L., MEGA, M., GRAY, K., ROSENBERG-THOMPSON, S., CARUSI, D. A. & GORNBEIN, J. 1994b. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology,* 44**,** 2308-14.

DARBY, M. M., YOLKEN, R. H. & SABUNCIYAN, S. 2016. Consistently altered expression of gene sets in postmortem brains of individuals with major psychiatric disorders. *Translational Psychiatry,* 6**,** e890.

DARREL A. REGIER, WILLIAM E. NARROW, DIANA E. CLARKE, HELENA C. KRAEMER, S. JANET KURAMOTO, EMILY A. KUHL & DAVID J. KUPFER 2013. DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses. *American Journal of Psychiatry,* 170**,** 59-70.

DE BERARDIS, D., MARINI, S., FORNARO, M., SRINIVASAN, V., IASEVOLI, F., TOMASETTI, C., VALCHERA, A., PERNA, G., QUERA-SALVA, M. A., MARTINOTTI, G. & DI GIANNANTONIO, M. 2013. The melatonergic system in mood and anxiety disorders and the role of agomelatine: implications for clinical practice. *Int J Mol Sci,* 14**,** 12458-83.

DE JONG, S., NEWHOUSE, S. J., PATEL, H., LEE, S., DEMPSTER, D., CURTIS, C., PAYA-CANO, J., MURPHY, D., WILSON, C. E., HORDER, J., MENDEZ, M. A., ASHERSON, P., RIVERA, M., COSTELLO, H., MALTEZOS, S., WHITWELL, S., PITTS, M., TYE, C., ASHWOOD, K. L., BOLTON, P., CURRAN, S., MCGUFFIN, P., DOBSON, R. & BREEN, G. 2016. Immune signatures and disorder-specific patterns in a cross-disorder gene expression analysis. *The British Journal of Psychiatry,* 209**,** 202-208.

DÍAZ-URIARTE, R. & ALVAREZ DE ANDRÉS, S. 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics,* 7**,** 3-3.

DIMOPOULOS, N., PIPERI, C., SALONICIOTI, A., MITSONIS, C., LIAPPAS, I., LEA, R. W. & KALOFOUTIS, A. 2006. Elevation of plasma concentration of adhesion

molecules in late-life depression. *International Journal of Geriatric Psychiatry,* 21, 965-971.

DORIS, A., EBMEIER, K. & SHAJAHAN, P. 1999. Depressive illness. *The Lancet,* 354, 1369-1375.

DOTSON, V. M., DAVATZIKOS, C., KRAUT, M. A. & RESNICK, S. M. 2009. Depressive symptoms and brain volumes in older adults: a longitudinal magnetic resonance imaging study. *Journal of Psychiatry & Neuroscience : JPN,* 34, 367-375.

DRAGO, A., DE RONCHI, D. & SERRETTI, A. 2007. Incomplete Coverage of Candidate Genes: A Poorly Considered Bias. *Current Genomics,* 8, 476-483.

DUMAN, R. S., MALBERG, J. & THOME, J. 1999. Neural plasticity to stress and antidepressant treatment. *Biol Psychiatry,* 46, 1181-91.

DURIC, V., BANASR, M., STOCKMEIER, C. A., SIMEN, A. A., NEWTON, S. S., OVERHOLSER, J. C., JURJUS, G. J., DIETER, L. & DUMAN, R. S. 2013. Altered expression of synapse and glutamate related genes in post-mortem hippocampus of depressed subjects. *Int J Neuropsychopharmacol,* 16, 69-82.

ERIKSSON, P. S., PERFILIEVA, E., BJORK-ERIKSSON, T., ALBORN, A. M., NORDBORG, C., PETERSON, D. A. & GAGE, F. H. 1998. Neurogenesis in the adult human hippocampus. *Nat Med,* 4, 1313-7.

EVANS, D. M., VISSCHER, P. M. & WRAY, N. R. 2009. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Human Molecular Genetics,* 18, 3525-3531.

EVANS, S. J., CHOUDARY, P. V., NEAL, C. R., LI, J. Z., VAWTER, M. P., TOMITA, H., LOPEZ, J. F., THOMPSON, R. C., MENG, F., STEAD, J. D., WALSH, D. M., MYERS, R. M., BUNNEY, W. E., WATSON, S. J., JONES, E. G. & AKIL, H. 2004. Dysregulation of the fibroblast growth factor system in major depression. *Proceedings of the National Academy of Sciences of the United States of America,* 101, 15506-15511.

EYRE, H. A., AIR, T., PRADHAN, A., JOHNSTON, J., LAVRETSKY, H., STUART, M. J. & BAUNE, B. T. 2016. A meta-analysis of chemokines in major depression. *Progress in neuro-psychopharmacology & biological psychiatry,* 68, 1-8.

FENG, Y., KAPORNAI, K., KISS, E., TAMÁS, Z., MAYER, L., BAJI, I., DARÓCZI, G., BENÁK, I., KOTHENCNÉ, V. O., DOMBOVÁRI, E., KACZVINSZK, E., BESNYŐ, M., GÁDOROS, J., SZÉKELY, J., KOVACS, M., VETRÓ, Á., KENNEDY, J. L. & BARR, C. L. 2010. Association of the GABRD Gene and Childhood-Onset Mood Disorders. *Genes, brain, and behavior,* 9, 668-672.

FERRARI, A. J., CHARLSON, F. J., NORMAN, R. E., PATTEN, S. B., FREEDMAN, G., MURRAY, C. J. L., VOS, T. & WHITEFORD, H. A. 2013. Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010. *PLOS Medicine,* 10, e1001547.

FIANNACA, A., LA ROSA, M., LA PAGLIA, L., RIZZO, R. & URSO, A. 2015. Analysis of miRNA expression profiles in breast cancer using biclustering. *BMC Bioinformatics,* 16, S7-S7.

FISKE, A., WETHERELL, J. L. & GATZ, M. 2009. Depression in Older Adults. *Annual Review of Clinical Psychology,* 5, 363-389.

FLINT, J. & KENDLER, KENNETH S. 2014. The Genetics of Major Depression. *Neuron,* 81, 1214.

FOLEY, D. L., MORLEY, K. I., MADDEN, P. A. F., HEATH, A. C., WHITFIELD, J. B. & MARTIN, N. G. 2010. Major depression and the metabolic syndrome. *Twin research and human genetics : the official journal of the International Society for Twin Studies,* 13, 347-358.

FOLSTEIN, M. F., FOLSTEIN, S. E. & MCHUGH, P. R. 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res,* 12**,** 189-98.

FREDRICKSON, B. L., GREWEN, K. M., COFFEY, K. A., ALGOE, S. B., FIRESTINE, A. M., AREVALO, J. M., MA, J. & COLE, S. W. 2013. A functional genomic perspective on human well-being. *Proc Natl Acad Sci U S A,* 110**,** 13684-9.

FREIS, E. D. 1954. Mental Depression in Hypertensive Patients Treated for Long Periods with Large Doses of Reserpine. *New England Journal of Medicine,* 251**,** 1006-1008.

FRIES, G. R., QUEVEDO, J., ZENI, C. P., KAZIMI, I. F., ZUNTA-SOARES, G., SPIKER, D. E., BOWDEN, C. L., WALSS-BASS, C. & SOARES, J. C. 2017. Integrated transcriptome and methylome analysis in youth at high risk for bipolar disorder: a preliminary analysis. *Translational Psychiatry,* 7**,** e1059.

GAITERI, C., DING, Y., FRENCH, B., TSENG, G. C. & SIBILLE, E. 2014. Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes Brain Behav,* 13**,** 13-24.

GAITERI, C., GUILLOUX, J.-P., LEWIS, D. A. & SIBILLE, E. 2010a. Altered Gene Synchrony Suggests a Combined Hormone-Mediated Dysregulated State in Major Depression. *PLoS ONE,* 5**,** e9970.

GAITERI, C., GUILLOUX, J. P., LEWIS, D. A. & SIBILLE, E. 2010b. Altered gene synchrony suggests a combined hormone-mediated dysregulated state in major depression. *PLoS ONE,* 5.

GANDAL, M. J., HANEY, J., PARIKSHAK, N., LEPPA, V., HORVATH, S. & GESCHWIND, D. H. 2016. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *bioRxiv*.

GARBETT, K. A., VERECZKEI, A., KALMAN, S., BROWN, J. A., TAYLOR, W. D., FALUDI, G., KORADE, Z., SHELTON, R. C. & MIRNICS, K. 2015a. Coordinated messenger RNA/microRNA changes in fibroblasts of patients with major depression. *Biol Psychiatry,* 77**,** 256-65.

GARBETT, K. A., VERECZKEI, A., KÁLMÁN, S., BROWN, J. A., TAYLOR, W. D., FALUDI, G., KORADE, Z., SHELTON, R. C. & MIRNICS, K. 2015b. Coordinated messenger RNA/microRNA changes in fibroblasts of patients with major depression. *Biological Psychiatry,* 77**,** 256-265.

GBD 2017. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet,* 390**,** 1260-1344.

GHAHRAMANI, Z. 2004. Unsupervised Learning. *In:* BOUSQUET, O., VON LUXBURG, U. & RÄTSCH, G. (eds.) *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures.* Berlin, Heidelberg: Springer Berlin Heidelberg.

GIBNEY, E. R. & NOLAN, C. M. 2010. Epigenetics and gene expression. *Heredity,* 105**,** 4-13.

GONG, Q. & HE, Y. 2015. Depression, Neuroimaging and Connectomics: A Selective Overview. *Biological Psychiatry,* 77**,** 223-235.

GORENSHTEYN, D., ZASLAVSKY, E., FRIBOURG, M., PARK, CHRISTOPHER Y., WONG, AARON K., TADYCH, A., HARTMANN, BORIS M., ALBRECHT, RANDY A., GARCÍA-SASTRE, A., KLEINSTEIN, STEVEN H., TROYANSKAYA, OLGA G. & SEALFON, STUART C. 2015. Interactive Big Data Resource to Elucidate Human Immune Pathways and Diseases. *Immunity,* 43**,** 605-614.

GREENBERG, P. E., FOURNIER, A.-A., SISITSKY, T., PIKE, C. T. & KESSLER, R. C. 2015. The Economic Burden of Adults With Major Depressive Disorder in the United States (2005 and 2010). *J Clin Psychiatry*

GRIEVE, S. M., KORGAONKAR, M. S., KOSLOW, S. H., GORDON, E. & WILLIAMS, L. M. 2013. Widespread reductions in gray matter volume in depression. *NeuroImage: Clinical,* 3, 332-339.

GUILLOUX, J.-P., BASSI, S., DING, Y., WALSH, C., TURECKI, G., TSENG, G., CYRANOWSKI, J. M. & SIBILLE, E. 2014. Testing the Predictive Value of Peripheral Gene Expression for Nonremission Following Citalopram Treatment for Major Depression. *Neuropsychopharmacology,* 40, 701.

GUPTA, R., DEWAN, I., BHARTI, R. & BHATTACHARYA, A. 2012. Differential Expression Analysis for RNA-Seq Data. *ISRN Bioinformatics,* 2012, 8.

HAAPAKOSKI, R., EBMEIER, K. P., ALENIUS, H. & KIVIMÄKI, M. 2016. Innate and adaptive immunity in the development of depression: An update on current knowledge and technological advances. *Progress in Neuro-Psychopharmacology and Biological Psychiatry,* 66, 63-72.

HALARIS, A. 2017. Inflammation-Associated Co-morbidity Between Depression and Cardiovascular Disease. *Curr Top Behav Neurosci,* 31, 45-70.

HARDEVELD, F., SPIJKER, J., VREEBURG, S. A., GRAAF, R. D., HENDRIKS, S. M., LICHT, C. M., NOLEN, W. A., PENNINX, B. W. & BEEKMAN, A. T. 2014. Increased cortisol awakening response was associated with time to recurrence of major depressive disorder. *Psychoneuroendocrinology,* 50, 62-71.

HARRISON, R. N. S., MURRAY, R. M., LEE, S. H., PAYA CANO, J., DEMPSTER, D., CURTIS, C. J., DIMA, D., GAUGHRAN, F., BREEN, G. & DE JONG, S. 2016. Gene-expression analysis of clozapine treatment in whole blood of patients with psychosis. *Psychiatric Genetics,* 26, 211-217.

HASLER, G. 2010. PATHOPHYSIOLOGY OF DEPRESSION: DO WE HAVE ANY SOLID EVIDENCE. *World Psychiatry,* 9, 155-161.

HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer.

HAY, S. I., ABAJOBIR, A. A., ABATE, K. H., ABBAFATI, C., ABBAS, K. M., ABD-ALLAH, F., ABDULKADER, R. S., ABDULLE, A. M., ABEBO, T. A., ABERA, S. F., ABOYANS, V., ABU-RADDAD, L. J., ACKERMAN, I. N., ADEDEJI, I. A., ADETOKUNBOH, O., AFSHIN, A., AGGARWAL, R., AGRAWAL, S., AGRAWAL, A., AHMED, M. B., AICHOUR, M. T. E., AICHOUR, A. N., AICHOUR, I., AIYAR, S., AKINYEMIJU, T. F., AKSEER, N., AL LAMI, F. H., ALAHDAB, F., AL-ALY, Z., ALAM, K., ALAM, N., ALAM, T., ALASFOOR, D., ALENE, K. A., ALI, R., ALIZADEH-NAVAEI, R., ALKAABI, J. M., ALKERWI, A. A., ALLA, F., ALLEBECK, P., ALLEN, C., AL-MASKARI, F., ALMAZROA, M. A., AL-RADDADI, R., ALSHARIF, U., ALSOWAIDI, S., ALTHOUSE, B. M., ALTIRKAWI, K. A., ALVIS-GUZMAN, N., AMARE, A. T., AMINI, E., AMMAR, W., AMOAKO, Y. A., ANSHA, M. G., ANTONIO, C. A. T., ANWARI, P., ÄRNLÖV, J., ARORA, M., ARTAMAN, A., ARYAL, K. K., ASGEDOM, S. W., ATEY, T. M., ATNAFU, N. T., AVILA-BURGOS, L., AVOKPAHO, E. F. G. A., AWASTHI, A., AWASTHI, S., AZARPAZHOOH, M. R., AZZOPARDI, P., BABALOLA, T. K., BACHA, U., BADAWI, A., BALAKRISHNAN, K., BANNICK, M. S., BARAC, A., BARKER-COLLO, S. L., BÄRNIGHAUSEN, T., BARQUERA, S., BARRERO, L. H., BASU, S., BATTISTA, R., BATTLE, K. E., BAUNE, B. T., BAZARGAN-HEJAZI, S., BEARDSLEY, J., BEDI, N., BÉJOT, Y., BEKELE, B. B., BELL, M. L., BENNETT, D. A., BENNETT, J. R., BENSENOR, I. M., BENSON, J., BERHANE, A., BERHE, D. F., BERNABÉ, E., BETSU, B. D., BEURAN, M., BEYENE, A. S., BHANSALI, A., et al. 2017. Global, regional, and national disability-adjusted life-years (DALYs)

for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet,* 390**,** 1260-1344.

HEK, K., DEMIRKAN, A., LAHTI, J., TERRACCIANO, A., TEUMER, A., CORNELIS, M. C., AMIN, N., BAKSHIS, E., BAUMERT, J., DING, J., LIU, Y., MARCIANTE, K., MEIRELLES, O., NALLS, M. A., SUN, Y. V., VOGELZANGS, N., YU, L., BANDINELLI, S., BENJAMIN, E. J., BENNETT, D. A., BOOMSMA, D., CANNAS, A., COKER, L. H., DE GEUS, E., DE JAGER, P. L., DIEZ-ROUX, A. V., PURCELL, S., HU, F. B., RIMMA, E. B., HUNTER, D. J., JENSEN, M. K., CURHAN, G., RICE, K., PENMAN, A. D., ROTTER, J. I., SOTOODEHNIA, N., EMENY, R., ERIKSSON, J. G., EVANS, D. A., FERRUCCI, L., FORNAGE, M., GUDNASON, V., HOFMAN, A., ILLIG, T., KARDIA, S., KELLY-HAYES, M., KOENEN, K., KRAFT, P., KUNINGAS, M., MASSARO, J. M., MELZER, D., MULAS, A., MULDER, C. L., MURRAY, A., OOSTRA, B. A., PALOTIE, A., PENNINX, B., PETERSMANN, A., PILLING, L. C., PSATY, B., RAWAL, R., REIMAN, E. M., SCHULZ, A., SHULMAN, J. M., SINGLETON, A. B., SMITH, A. V., SUTIN, A. R., UITTERLINDEN, A. G., VOLZKE, H., WIDEN, E., YAFFE, K., ZONDERMAN, A. B., CUCCA, F., HARRIS, T., LADWIG, K. H., LLEWELLYN, D. J., RAIKKONEN, K., TANAKA, T., VAN DUIJN, C. M., GRABE, H. J., LAUNER, L. J., LUNETTA, K. L., MOSLEY, T. H., JR., NEWMAN, A. B., TIEMEIER, H. & MURABITO, J. 2013. A genome-wide association study of depressive symptoms. *Biol Psychiatry,* 73**,** 667-78.

HIRUNSATIT, R., GEORGE, E. D., LIPSKA, B. K., ELWAFI, H. M., SANDER, L., YRIGOLLEN, C. M., GELERNTER, J., GRIGORENKO, E. L., LAPPALAINEN, J., MANE, S., NAIRN, A. C., KLEINMAN, J. E. & SIMEN, A. A. 2009. Twenty-one-base-pair insertion polymorphism creates an enhancer element and potentiates SLC6A1 GABA transporter promoter activity. *Pharmacogenet Genomics,* 19**,** 53-65.

HSIAO, T.-H., CHIU, Y.-C., HSU, P.-Y., LU, T.-P., LAI, L.-C., TSAI, M.-H., HUANG, T. H. M., CHUANG, E. Y. & CHEN, Y. 2016. Differential network analysis reveals the genome-wide landscape of estrogen receptor modulation in hormonal cancers. *Scientific Reports,* 6**,** 23035.

HSU, C.-L., JUAN, H.-F. & HUANG, H.-C. 2015. Functional Analysis and Characterization of Differential Coexpression Networks. *Scientific Reports,* 5**,** 13295.

HYDE, C. L., NAGLE, M. W., TIAN, C., CHEN, X., PACIGA, S. A. & WENDLAND, J. R. 2016. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. 48**,** 1031-6.

INIESTA, R., STAHL, D. & MCGUFFIN, P. 2016. Machine learning, statistical learning and the future of biological research in psychiatry. *Psychological Medicine,* 46**,** 2455-2465.

IWAMOTO, K., KAKIUCHI, C., BUNDO, M., IKEDA, K. & KATO, T. 2004. Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. *Mol Psychiatry,* 9**,** 406-16.

JAKOBSDOTTIR, J., GORIN, M. B., CONLEY, Y. P., FERRELL, R. E. & WEEKS, D. E. 2009. Interpretation of Genetic Association Studies: Markers with Replicated Highly Significant Odds Ratios May Be Poor Classifiers. *PLOS Genetics,* 5**,** e1000337.

JANSEN, R., BATISTA, S., BROOKS, A. I., TISCHFIELD, J. A., WILLEMSEN, G., VAN GROOTHEEST, G., HOTTENGA, J.-J., MILANESCHI, Y., MBAREK, H., MADAR, V., PEYROT, W., VINK, J. M., VERWEIJ, C. L., DE GEUS, E. J., SMIT, J. H., WRIGHT, F. A., SULLIVAN, P. F., BOOMSMA, D. I. & PENNINX, B. W. 2014. Sex differences in the human peripheral blood transcriptome. *BMC Genomics,* 15**,** 33.

JANSEN, R., PENNINX, B. W., MADAR, V. & XIA, K. 2016. Gene expression in major depressive disorder. 21, 339-47.

JIANG, M., QIN, P. & YANG, X. 2014. Comorbidity between depression and asthma via immune-inflammatory pathways: A meta-analysis. *Journal of Affective Disorders,* 166, 22-29.

KALMAN, S., GARBETT, K. A., JANKA, Z. & MIRNICS, K. 2016. Human dermal fibroblasts in psychiatry research. *Neuroscience,* 320, 105-121.

KANG, H. J., ADAMS, D. H., SIMEN, A., SIMEN, B. B., RAJKOWSKA, G., STOCKMEIER, C. A., OVERHOLSER, J. C., MELTZER, H. Y., JURJUS, G. J., KONICK, L. C., NEWTON, S. S. & DUMAN, R. S. 2007. Gene expression profiling in postmortem prefrontal cortex of major depressive disorder. *J Neurosci,* 27, 13329-40.

KAWABATA, H. 2018. Transferrin and transferrin receptors update. *Free Radical Biology and Medicine*.

KHUMALO, H., GOMO, Z. A. R., MOYO, V. M., GORDEUK, V. R., SAUNGWEME, T., ROUAULT, T. A. & GANGAIDZO, I. T. 1998. Serum transferrin receptors are decreased in the presence of iron overload. *Clinical Chemistry,* 44, 40-44.

KIM, S., HWANG, Y., WEBSTER, M. J. & LEE, D. 2016. Differential activation of immune/inflammatory response-related co-expression modules in the hippocampus across the major psychiatric disorders. *Molecular Psychiatry,* 21, 376-385.

KIRSCH, I. 2014. Antidepressants and the Placebo Effect. *Z Psychol,* 222, 128-134.

KLEMPAN, T. A., SEQUEIRA, A., CANETTI, L., LALOVIC, A., ERNST, C., FFRENCH-MULLEN, J. & TURECKI, G. 2009. Altered expression of genes involved in ATP biosynthesis and GABAergic neurotransmission in the ventral prefrontal cortex of suicides with and without major depression. *Mol Psychiatry,* 14, 175-89.

KOHEN, R., DOBRA, A., TRACY, J. H. & HAUGEN, E. 2014a. Transcriptome profiling of human hippocampus dentate gyrus granule cells in mental illness. *Transl Psychiatry,* 4, e366.

KOHEN, R., DOBRA, A., TRACY, J. H. & HAUGEN, E. 2014b. Transcriptome profiling of human hippocampus dentate gyrus granule cells in mental illness. *Translational Psychiatry,* 4.

KOHLER-FORSBERG, O., BUTTENSCHON, H. N., TANSEY, K. E., MAIER, W., HAUSER, J., DERNOVSEK, M. Z., HENIGSBERG, N., SOUERY, D., FARMER, A., RIETSCHEL, M., MCGUFFIN, P., AITCHISON, K. J., UHER, R. & MORS, O. 2017. Association between C-reactive protein (CRP) with depression symptom severity and specific depressive symptoms in major depression. *Brain Behav Immun,* 62, 344-350.

KÖHLER, O., BENROS, M. E., NORDENTOFT, M. & ET AL. 2014. Effect of anti-inflammatory treatment on depression, depressive symptoms, and adverse effects: A systematic review and meta-analysis of randomized clinical trials. *JAMA Psychiatry,* 71, 1381-1391.

KÖHLER, O., KROGH, J., MORS, O. & BENROS, M. E. 2016. Inflammation in Depression and the Potential for Anti-Inflammatory Treatment. *Current Neuropharmacology,* 14, 732-742.

KROENKE K & R., S. 2002. The PHQ-9: A New Depression Diagnostic and Severity Measure. *Psychiatr Ann*.

KROENKE, K., SPITZER, R. L. & WILLIAMS, J. B. 2001. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med,* 16, 606-13.

KUBOTA, T., MIYAKE, K. & HIRASAWA, T. 2012. Epigenetic understanding of gene-environment interactions in psychiatric disorders: a new concept of clinical genetics. *Clinical Epigenetics,* 4, 1.

LAI, C. K. Y. 2014. The merits and problems of Neuropsychiatric Inventory as an assessment tool in people with dementia and other neurological disorders. *Clinical Interventions in Aging,* 9, 1051-1061.

LANDIS, J. R. & KOCH, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics,* 33, 159-174.

LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics,* 9, 559.

LANGFELDER, P. & HORVATH, S. 2012. Fast R Functions for Robust Correlations and Hierarchical Clustering. *Journal of statistical software,* 46, i11.

LAREAU, C. A., WHITE, B. C., OBERG, A. L. & MCKINNEY, B. A. 2015. Differential co-expression network centrality and machine learning feature selection for identifying susceptibility hubs in networks with scale-free structure. *BioData Mining,* 8, 5.

LAU, C., NYGÅRD, S., FURE, H., OLSTAD, O. K., HOLDEN, M., LAPPEGÅRD, K. T., BREKKE, O.-L., ESPEVIK, T., HOVIG, E. & MOLLNES, T. E. 2015. CD14 and Complement Crosstalk and Largely Mediate the Transcriptional Response to Escherichia coli in Human Whole Blood as Revealed by DNA Microarray. *PLOS ONE,* 10, e0117261.

LEDAY, G. G. R., VERTES, P. E., RICHARDSON, S., GREENE, J. R., REGAN, T., KHAN, S., HENDERSON, R., FREEMAN, T. C., PARIANTE, C. M., HARRISON, N. A., PERRY, V. H., DREVETS, W. C., WITTENBERG, G. M. & BULLMORE, E. T. 2018. Replicable and Coupled Changes in Innate and Adaptive Immune Gene Expression in Two Case-Control Studies of Blood Microarrays in Major Depressive Disorder. *Biol Psychiatry,* 83, 70-80.

LEE, S. H., RIPKE, S., NEALE, B. M. & FARAONE, S. V. 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet,* 45, 984-994.

LEEK, J. T., JOHNSON, W. E., PARKER, H. S., JAFFE, A. E. & STOREY, J. D. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics,* 28, 882-3.

LEONTJEVAS, R., VAN HOOREN, S. & MULDERS, A. 2009. The Montgomery-Asberg Depression Rating Scale and the Cornell Scale for Depression in Dementia: a validation study with patients exhibiting early-onset dementia. *Am J Geriatr Psychiatry,* 17, 56-64.

LESPÉRANCE, F., FRASURE-SMITH, N., THÉROUX, P. & IRWIN, M. 2004. The Association Between Major Depression and Levels of Soluble Intercellular Adhesion Molecule 1, Interleukin-6, and C-Reactive Protein in Patients With Recent Acute Coronary Syndromes. *American Journal of Psychiatry,* 161, 271-277.

LEVINSON, D. F., MOSTAFAVI, S., MILANESCHI, Y., RIVERA, M., RIPKE, S., WRAY, N. R. & SULLIVAN, P. F. 2014. Genetic studies of major depressive disorder: Why are there no GWAS findings, and what can we do about it? *Biological psychiatry,* 76, 510-512.

LI, J. Z., BUNNEY, B. G., MENG, F., HAGENAUER, M. H., WALSH, D. M., VAWTER, M. P., EVANS, S. J., CHOUDARY, P. V., CARTAGENA, P., BARCHAS, J. D., SCHATZBERG, A. F., JONES, E. G., MYERS, R. M., WATSON, S. J., AKIL, H. & BUNNEY, W. E. 2013. Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. *Proceedings of the National Academy of Sciences,* 110, 9950-9955.

LI, M., KOUZMINA, E., MCCUSKER, M., RODIN, D., BOUTROS, P. C., PAIGE, C. J. & RODIN, G. 2017. Cytokines and depression in cancer patients and caregivers. *Neuropsychiatric Disease and Treatment,* 13, 2903-2911.

LIEW, C. C., MA, J., TANG, H. C., ZHENG, R. & DEMPSEY, A. A. 2006. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J Lab Clin Med,* 147**,** 126-32.

LIGUORI, I., RUSSO, G., CURCIO, F., BULLI, G., ARAN, L., DELLA-MORTE, D., GARGIULO, G., TESTA, G., CACCIATORE, F., BONADUCE, D. & ABETE, P. 2018. Oxidative stress, aging, and diseases. *Clinical interventions in aging,* 13**,** 757-772.

LIN, E. & TSAI, S. J. 2016. Genome-wide microarray analysis of gene expression profiling in major depression and antidepressant therapy. *Prog Neuropsychopharmacol Biol Psychiatry,* 64**,** 334-40.

LIN, S. M., DU, P., HUBER, W. & KIBBE, W. A. 2008. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Research,* 36**,** e11.

LIU, H. H., HU, Y., ZHENG, M., SUHOSKI, M. M., ENGLEMAN, E. G., DILL, D. L., HUDNALL, M., WANG, J., SPOLSKI, R., LEONARD, W. J. & PELTZ, G. 2012. Cd14 SNPs regulate the innate immune response. *Mol Immunol,* 51**,** 112-27.

LIU, W., GE, T., LENG, Y., PAN, Z., FAN, J., YANG, W. & CUI, R. 2017. The Role of Neural Plasticity in Depression: From Hippocampus to Prefrontal Cortex. *Neural Plasticity,* 2017**,** 6871089.

LIU, Z., LI, X., SUN, N., XU, Y., MENG, Y., YANG, C., WANG, Y. & ZHANG, K. 2014. Microarray profiling and co-expression network analysis of circulating lncRNAs and mRNAs associated with major depressive disorder. *PLoS One,* 9**,** e93388.

LOFTIS, J. M., HUCKANS, M. & MORASCO, B. J. 2010. Neuroimmune mechanisms of cytokine-induced depression: Current theories and novel treatment strategies. *Neurobiology of disease,* 37**,** 519-533.

LOPRESTI, A. L., MAKER, G. L., HOOD, S. D. & DRUMMOND, P. D. 2014. A review of peripheral biomarkers in major depression: the potential of inflammatory and oxidative stress biomarkers. *Prog Neuropsychopharmacol Biol Psychiatry,* 48**,** 102-11.

LORENZETTI, V., ALLEN, N. B., FORNITO, A. & YUCEL, M. 2009. Structural brain abnormalities in major depressive disorder: a selective review of recent MRI studies. *J Affect Disord,* 117**,** 1-17.

LOTAN, A., FENCKOVA, M., BRALTEN, J., ALTTOA, A., DIXSON, L., WILLIAMS, R. W. & VAN DER VOET, M. 2014. Neuroinformatic analyses of common and distinct genetic components associated with major neuropsychiatric disorders. *Frontiers in Neuroscience,* 8**,** 331.

LU, J. T., MUCHIR, A., NAGY, P. L. & WORMAN, H. J. 2011. LMNA cardiomyopathy: cell biology and genetics meet clinical medicine. *Disease Models & Mechanisms,* 4**,** 562-568.

LU, N. Z. & CIDLOWSKI, J. A. 2005. Translational regulatory mechanisms generate N-terminal glucocorticoid receptor isoforms with unique transcriptional target genes. *Mol Cell,* 18**,** 331-42.

LUSCHER, B. & FUCHS, T. 2015. GABAergic control of depression-related brain states. *Advances in pharmacology (San Diego, Calif.),* 73**,** 97-144.

MACHADO-VIEIRA, R., SALVADORE, G., LUCKENBAUGH, D. A., MANJI, H. K. & ZARATE, C. A., JR. 2008. Rapid onset of antidepressant action: a new paradigm in the research and treatment of major depressive disorder. *J Clin Psychiatry,* 69**,** 946-58.

MAES, M. 1999. Major depression and activation of the inflammatory response system. *Adv Exp Med Biol,* 461**,** 25-46.

MALKI, K., PAIN, O., TOSTO, M. G., DU RIETZ, E., CARBONI, L. & SCHALKWYK, L. C. 2015. Identification of genes and gene pathways associated with major depressive

disorder by integrative brain analysis of rat and human prefrontal cortex transcriptomes. *Transl Psychiatry,* 5**,** e519.

MALKI, K., TOSTO, M. G., JUMABHOY, I. & LOURDUSAMY, A. 2013. Integrative mouse and human mRNA studies using WGCNA nominates novel candidate genes involved in the pathogenesis of major depressive disorder. *Pharmacogenomics,* 14**,** 1979-1990.

MARIC, N. P. & ADZIC, M. 2013. Pharmacological modulation of HPA axis in depression - new avenues for potential therapeutic benefits. *Psychiatr Danub,* 25**,** 299-305.

MARONDE, E., SAADE, A., ACKERMANN, K., GOUBRAN-BOTROS, H., PAGAN, C., BUX, R., BOURGERON, T., DEHGHANI, F. & STEHLE, J. H. 2011. Dynamics in enzymatic protein complexes offer a novel principle for the regulation of melatonin synthesis in the human pineal gland. *J Pineal Res,* 51**,** 145-55.

MATHEW, S. J., KEEGAN, K. & SMITH, L. 2005. Glutamate modulators as novel interventions for mood disorders. *Rev Bras Psiquiatr,* 27**,** 243-8.

MCAFOOSE, J. & BAUNE, B. T. 2009. Evidence for a cytokine model of cognitive function. *Neurosci Biobehav Rev,* 33**,** 355-66.

MCCLUNG, C. A. 2007. Circadian Genes, Rhythms and the Biology of Mood Disorders. *Pharmacology & therapeutics,* 114**,** 222-232.

MEHTA, D., MENKE, A. & BINDER, E. B. 2010. Gene Expression Studies in Major Depression. *Current Psychiatry Reports,* 12**,** 135-144.

MENKE, A., ARLOTH, J., PUTZ, B., WEBER, P., KLENGEL, T., MEHTA, D., GONIK, M., REX-HAFFNER, M., RUBEL, J., UHR, M., LUCAE, S., DEUSSING, J. M., MULLER-MYHSOK, B., HOLSBOER, F. & BINDER, E. B. 2012. Dexamethasone stimulated gene expression in peripheral blood is a sensitive marker for glucocorticoid receptor resistance in depressed patients. *Neuropsychopharmacology,* 37**,** 1455-64.

METCALF, T. U., CUBAS, R. A., GHNEIM, K., CARTWRIGHT, M. J., GREVENYNGHE, J. V., RICHNER, J. M., OLAGNIER, D. P., WILKINSON, P. A., CAMERON, M. J., PARK, B. S., HISCOTT, J. B., DIAMOND, M. S., WERTHEIMER, A. M., NIKOLICH-ZUGICH, J. & HADDAD, E. K. 2015. Global analyses revealed age-related alterations in innate immune responses after stimulation of pathogen recognition receptors. *Aging Cell,* 14**,** 421-32.

MILLS, N. T., SCOTT, J. G., WRAY, N. R., COHEN-WOODS, S. & BAUNE, B. T. 2013. Research review: the role of cytokines in depression in adolescents: a systematic review. *J Child Psychol Psychiatry,* 54**,** 816-35.

MISTRY, M., GILLIS, J. & PAVLIDIS, P. 2013. Meta-analysis of gene coexpression networks in the post-mortem prefrontal cortex of patients with schizophrenia and unaffected controls. *BMC Neuroscience,* 14**,** 105-105.

MOSTAFAVI, S., BATTLE, A., ZHU, X., POTASH, J. B., WEISSMAN, M. M., SHI, J., BECKMAN, K., HAUDENSCHILD, C., MCCORMICK, C., MEI, R., GAMEROFF, M. J., GINDES, H., ADAMS, P., GOES, F. S., MONDIMORE, F. M., MACKINNON, D. F., NOTES, L., SCHWEIZER, B., FURMAN, D., MONTGOMERY, S. B., URBAN, A. E., KOLLER, D. & LEVINSON, D. F. 2014. Type I interferon signaling genes in recurrent major depression: increased expression detected by whole-blood RNA sequencing. *Mol Psychiatry,* 19**,** 1267-74.

NICODEMUS, K. K. & MALLEY, J. D. 2009. Predictor correlation impacts machine learning algorithms: implications for genomic studies. *Bioinformatics,* 25**,** 1884-90.

OKBAY, A., BASELMANS, B. M. L., DE NEVE, J.-E., TURLEY, P., NIVARD, M. G., FONTANA, M. A., MEDDENS, S. F. W., LINNÉR, R. K., RIETVELD, C. A., DERRINGER, J., GRATTEN, J., LEE, J. J., LIU, J. Z., DE VLAMING, R., AHLUWALIA, T. S., BUCHWALD, J., CAVADINO, A., FRAZIER-WOOD, A. C., FURLOTTE, N. A., GARFIELD, V., GEISEL, M. H., GONZALEZ, J. R., HAITJEMA,

S., KARLSSON, R., VAN DER LAAN, S. W., LADWIG, K.-H., LAHTI, J., VAN DER LEE, S. J., LIND, P. A., LIU, T., MATTESON, L., MIHAILOV, E., MILLER, M. B., MINICA, C. C., NOLTE, I. M., MOOK-KANAMORI, D., VAN DER MOST, P. J., OLDMEADOW, C., QIAN, Y., RAITAKARI, O., RAWAL, R., REALO, A., RUEEDI, R., SCHMIDT, B., SMITH, A. V., STERGIAKOULI, E., TANAKA, T., TAYLOR, K., THORLEIFSSON, G., WEDENOJA, J., WELLMANN, J., WESTRA, H.-J., WILLEMS, S. M., ZHAO, W., LIFELINES COHORT, S., AMIN, N., BAKSHI, A., BERGMANN, S., BJORNSDOTTIR, G., BOYLE, P. A., CHERNEY, S., COX, S. R., DAVIES, G., DAVIS, O. S. P., DING, J., DIREK, N., EIBICH, P., EMENY, R. T., FATEMIFAR, G., FAUL, J. D., FERRUCCI, L., FORSTNER, A. J., GIEGER, C., GUPTA, R., HARRIS, T. B., HARRIS, J. M., HOLLIDAY, E. G., HOTTENGA, J.-J., DE JAGER, P. L., KAAKINEN, M. A., KAJANTIE, E., KARHUNEN, V., KOLCIC, I., KUMARI, M., LAUNER, L. J., FRANKE, L., LI-GAO, R., LIEWALD, D. C., KOINI, M., LOUKOLA, A., MARQUES-VIDAL, P., MONTGOMERY, G. W., MOSING, M. A., PATERNOSTER, L., PATTIE, A., PETROVIC, K. E., PULKKI-RÅBACK, L., QUAYE, L., RÄIKKÖNEN, K., RUDAN, I., et al. 2016. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics,* 48**,** 624.

OKSER, S., PAHIKKALA, T., AIROLA, A., SALAKOSKI, T., RIPATTI, S. & AITTOKALLIO, T. 2014. Regularized Machine Learning in the Genetic Prediction of Complex Traits. *PLOS Genetics,* 10**,** e1004754.

PANDEY, G. N., RIZAVI, H. S., REN, X., FAREED, J., HOPPENSTEADT, D. A., ROBERTS, R. C., CONLEY, R. R. & DWIVEDI, Y. 2012. Proinflammatory cytokines in the prefrontal cortex of teenage suicide victims. *J Psychiatr Res,* 46**,** 57-63.

PAPAKOSTAS, G. I. A. F., MAURIZIO 2010. *Pharmacotherapy for Depression and Treatment-Resistant Depression*.

PARIANTE, C. M. & LIGHTMAN, S. L. 2008. The HPA axis in major depression: classical theories and new developments. *Trends in Neurosciences,* 31**,** 464-468.

PARIKSHAK, NEELROOP N., LUO, R., ZHANG, A., WON, H., LOWE, JENNIFER K., CHANDRAN, V., HORVATH, S. & GESCHWIND, DANIEL H. 2013. Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism. *Cell,* 155**,** 1008-1021.

PEARSON-FUHRHOP, K. M., DUNN, E. C., MORTERO, S., DEVAN, W. J., FALCONE, G. J., LEE, P., HOLMES, A. J., HOLLINSHEAD, M. O., ROFFMAN, J. L., SMOLLER, J. W., ROSAND, J. & CRAMER, S. C. 2014. Dopamine Genetic Risk Score Predicts Depressive Symptoms in Healthy Adults and Adults with Depression. *PLoS ONE,* 9**,** e93772.

PONTES, B., GIRÁLDEZ, R. & AGUILAR-RUIZ, J. S. 2015. Biclustering on expression data: A review. *Journal of Biomedical Informatics,* 57**,** 163-180.

PORTO, G. & DE SOUSA, M. 2007. Iron overload and immunity. *World journal of gastroenterology,* 13**,** 4707-4715.

POWER, R. A., TANSEY, K. E., BUTTENSCHØN, H. N., COHEN-WOODS, S., BIGDELI, T., HALL, L. S., KUTALIK, Z., LEE, S. H., RIPKE, S., STEINBERG, S., TEUMER, A., VIKTORIN, A., WRAY, N. R., AROLT, V., BAUNE, B. T., BOOMSMA, D. I., BØRGLUM, A. D., BYRNE, E. M., CASTELAO, E., CRADDOCK, N., CRAIG, I. W., DANNLOWSKI, U., DEARY, I. J., DEGENHARDT, F., FORSTNER, A. J., GORDON, S. D., GRABE, H. J., GROVE, J., HAMILTON, S. P., HAYWARD, C., HEATH, A. C., HOCKING, L. J., HOMUTH, G., HOTTENGA, J. J., KLOIBER, S., KROGH, J., LANDÉN, M., LANG, M., LEVINSON, D. F., LICHTENSTEIN, P., LUCAE, S., MACINTYRE, D. J., MADDEN, P., MAGNUSSON, P. K. E., MARTIN, N. G., MCINTOSH, A. M., MIDDELDORP, C. M., MILANESCHI, Y.,

MONTGOMERY, G. W., MORS, O., MÜLLER-MYHSOK, B., NYHOLT, D. R., OSKARSSON, H., OWEN, M. J., PADMANABHAN, S., PENNINX, B. W. J. H., PERGADIA, M. L., PORTEOUS, D. J., POTASH, J. B., PREISIG, M., RIVERA, M., SHI, J., SHYN, S. I., SIGURDSSON, E., SMIT, J. H., SMITH, B. H., STEFANSSON, H., STEFANSSON, K., STROHMAIER, J., SULLIVAN, P. F., THOMSON, P., THORGEIRSSON, T. E., VAN DER AUWERA, S., WEISSMAN, M. M., BREEN, G. & LEWIS, C. M. 2017. Genome-wide Association for Major Depression Through Age at Onset Stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. *Biological Psychiatry,* 81**,** 325-335.

RIPKE, S., WRAY, N. R., LEWIS, C. M., HAMILTON, S. P., WEISSMAN, M. M., BREEN, G., BYRNE, E. M., BLACKWOOD, D. H., BOOMSMA, D. I., CICHON, S., HEATH, A. C., HOLSBOER, F., LUCAE, S., MADDEN, P. A., MARTIN, N. G., MCGUFFIN, P., MUGLIA, P., NOETHEN, M. M., PENNINX, B. P., PERGADIA, M. L., POTASH, J. B., RIETSCHEL, M., LIN, D., MULLER-MYHSOK, B., SHI, J., STEINBERG, S., GRABE, H. J., LICHTENSTEIN, P., MAGNUSSON, P., PERLIS, R. H., PREISIG, M., SMOLLER, J. W., STEFANSSON, K., UHER, R., KUTALIK, Z., TANSEY, K. E., TEUMER, A., VIKTORIN, A., BARNES, M. R., BETTECKEN, T., BINDER, E. B., BREUER, R., CASTRO, V. M., CHURCHILL, S. E., CORYELL, W. H., CRADDOCK, N., CRAIG, I. W., CZAMARA, D., DE GEUS, E. J., DEGENHARDT, F., FARMER, A. E., FAVA, M., FRANK, J., GAINER, V. S., GALLAGHER, P. J., GORDON, S. D., GORYACHEV, S., GROSS, M., GUIPPONI, M., HENDERS, A. K., HERMS, S., HICKIE, I. B., HOEFELS, S., HOOGENDIJK, W., HOTTENGA, J. J., IOSIFESCU, D. V., ISING, M., JONES, I., JONES, L., JUNG-YING, T., KNOWLES, J. A., KOHANE, I. S., KOHLI, M. A., KORSZUN, A., LANDEN, M., LAWSON, W. B., LEWIS, G., MACINTYRE, D., MAIER, W., MATTHEISEN, M., MCGRATH, P. J., MCINTOSH, A., MCLEAN, A., MIDDELDORP, C. M., MIDDLETON, L., MONTGOMERY, G. M., MURPHY, S. N., NAUCK, M., NOLEN, W. A., NYHOLT, D. R., O'DONOVAN, M., OSKARSSON, H., PEDERSEN, N., SCHEFTNER, W. A., SCHULZ, A., SCHULZE, T. G., SHYN, S. I., SIGURDSSON, E., SLAGER, S. L., SMIT, J. H., et al. 2013. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry,* 18**,** 497-511.

RITCHIE, M. D., HOLZINGER, E. R., LI, R., PENDERGRASS, S. A. & KIM, D. 2015. Methods of integrating data to uncover genotype–phenotype interactions. *Nature Reviews Genetics,* 16**,** 85.

RIVERA, M. & MCGUFFIN, P. 2015. The successful search for genetic loci associated with depression. *Genome Medicine,* 7**,** 92.

ROSENBLAT, J. D., CHA, D. S., MANSUR, R. B. & MCINTYRE, R. S. 2014. Inflamed moods: a review of the interactions between inflammation and mood disorders. *Prog Neuropsychopharmacol Biol Psychiatry,* 53**,** 23-34.

ROSTOKER, G., GRIUNCELLI, M., LORIDON, C., MAGNA, T., MACHADO, G., DRAHI, G., DAHAN, H., JANKLEWICZ, P. & COHEN, Y. 2015. Reassessment of Iron Biomarkers for Prediction of Dialysis Iron Overload: An MRI Study. *PLOS ONE,* 10**,** e0132006.

ROYCHOWDHURY, S. & CHINNAIYAN, A. M. 2016. Translating cancer genomes and transcriptomes for precision oncology. *CA Cancer J Clin,* 66**,** 75-88.

SACHDEV, P. S., BRODATY, H., REPPERMUND, S., KOCHAN, N. A., TROLLOR, J. N., DRAPER, B., SLAVIN, M. J., CRAWFORD, J., KANG, K., BROE, G. A., MATHER, K. A. & LUX, O. 2010. The Sydney Memory and Ageing Study (MAS): methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70-90 years. *Int Psychogeriatr,* 22**,** 1248-64.

SACHDEV, P. S., LAMMEL, A., TROLLOR, J. N., LEE, T., WRIGHT, M. J., AMES, D., WEN, W., MARTIN, N. G., BRODATY, H. & SCHOFIELD, P. R. 2012. A Comprehensive Neuropsychiatric Study of Elderly Twins: The Older Australian Twins Study. *Twin Research and Human Genetics,* 12**,** 573-582.

SAHAY, B., PATSEY, R. L., EGGERS, C. H., SALAZAR, J. C., RADOLF, J. D. & SELLATI, T. J. 2009. CD14 Signaling Restrains Chronic Inflammation through Induction of p38-MAPK/SOCS-Dependent Tolerance. *PLOS Pathogens,* 5**,** e1000687.

SANACORA, G., TRECCANI, G. & POPOLI, M. 2012. Towards a glutamate hypothesis of depression: An emerging frontier of neuropsychopharmacology for mood disorders. *Neuropharmacology,* 62**,** 63-77.

SANDERS, A. R., DRIGALENKO, E. I., DUAN, J., MOY, W., FREDA, J., GÖRING, H. H. H. & GEJMAN, P. V. 2017. Transcriptome sequencing study implicates immune-related genes differentially expressed in schizophrenia: new data and a meta-analysis. *Translational Psychiatry,* 7**,** e1093.

SAVITZ, J., FRANK, M. B., VICTOR, T., BEBAK, M., MARINO, J. H., BELLGOWAN, P. S., MCKINNEY, B. A., BODURKA, J., KENT TEAGUE, T. & DREVETS, W. C. 2013. Inflammation and neurological disease-related genes are differentially expressed in depressed patients with mood disorders and correlate with morphometric and functional imaging abnormalities. *Brain Behav Immun,* 31**,** 161-71.

SCHACHTER, J., MARTEL, J., LIN, C. S., CHANG, C. J., WU, T. R., LU, C. C., KO, Y. F., LAI, H. C., OJCIUS, D. M. & YOUNG, J. D. 2018. Effects of obesity on depression: A role for inflammation and the gut microbiota. *Brain Behav Immun,* 69**,** 1-8.

SCHATZBERG, A. F., KELLER, J., TENNAKOON, L., LEMBKE, A., WILLIAMS, G., KRAEMER, F. B., SARGINSON, J. E., LAZZERONI, L. C. & MURPHY, G. M. 2014. HPA axis genetic variation, cortisol and psychosis in major depression. *Mol Psychiatry,* 19**,** 220-227.

SCHILDKRAUD, J. J. 1965. THE CATECHOLAMINE HYPOTHESIS OF AFFECTIVE DISORDERS: A REVIEW OF SUPPORTING EVIDENCE. *American Journal of Psychiatry,* 122**,** 509-522.

SCHOEPP, D. D. 2001. Unveiling the functions of presynaptic metabotropic glutamate receptors in the central nervous system. *J Pharmacol Exp Ther,* 299**,** 12-20.

SEGMAN, R. H., GOLTSER-DUBNER, T., WEINER, I., CANETTI, L., GALILI-WEISSTUB, E., MILWIDSKY, A., PABLOV, V., FRIEDMAN, N. & HOCHNER-CELNIKIER, D. 2010. Blood mononuclear cell gene expression signature of postpartum depression. *Mol Psychiatry,* 15**,** 93-100, 2.

SEIFUDDIN, F., PIROOZNIA, M., JUDY, J. T., GOES, F. S., POTASH, J. B. & ZANDI, P. P. 2013. Systematic review of genome-wide gene expression studies of bipolar disorder. *BMC Psychiatry,* 13**,** 213-213.

SEQUEIRA, A., GWADRY, F. G., FFRENCH-MULLEN, J. M., CANETTI, L., GINGRAS, Y., CASERO, R. A., JR., ROULEAU, G., BENKELFAT, C. & TURECKI, G. 2006. Implication of SSAT by gene expression and genetic variation in suicide and major depression. *Arch Gen Psychiatry,* 63**,** 35-48.

SEQUEIRA, A., KLEMPAN, T., CANETTI, L., FFRENCH-MULLEN, J., BENKELFAT, C., ROULEAU, G. A. & TURECKI, G. 2007. Patterns of gene expression in the limbic system of suicides with and without major depression. *Mol Psychiatry,* 12**,** 640-55.

SEQUEIRA, A., MAMDANI, F., ERNST, C., VAWTER, M. P., BUNNEY, W. E., LEBEL, V., REHAL, S., KLEMPAN, T., GRATTON, A., BENKELFAT, C., ROULEAU, G. A., MECHAWAR, N. & TURECKI, G. 2009. Global brain gene expression analysis

links glutamatergic and GABAergic alterations to suicide and major depression. *PLoS One,* 4**,** e6585.

SERAFINI, G. 2012. Neuroplasticity and major depression, the role of modern antidepressant drugs. *World Journal of Psychiatry,* 2**,** 49-57.

SHEEHAN, D. V., LECRUBIER, Y., SHEEHAN, K. H., AMORIM, P., JANAVS, J., WEILLER, E., HERGUETA, T., BAKER, R. & DUNBAR, G. C. 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry,* 59**,** 22-33.

SHELTON, R. C., CLAIBORNE, J., SIDORYK-WEGRZYNOWICZ, M., REDDY, R., ASCHNER, M., LEWIS, D. A. & MIRNICS, K. 2011. Altered expression of genes involved in inflammation and apoptosis in frontal cortex in major depression. *Molecular Psychiatry,* 16**,** 751-762.

SIBILLE, E., ARANGO, V., GALFALVY, H. C., PAVLIDIS, P., ERRAJI-BENCHEKROUN, L., ELLIS, S. P. & JOHN MANN, J. 2004. Gene expression profiling of depression and suicide in human prefrontal cortex. *Neuropsychopharmacology,* 29**,** 351-61.

SIBILLE, E. & FRENCH, B. 2013. Biological substrates underpinning diagnosis of major depression. *International Journal of Neuropsychopharmacology,* 16**,** 1893-1909.

SIBILLE, E., WANG, Y., JOEYEN-WALDORF, J., GAITERI, C., SURGET, A., OH, S., BELZUNG, C., TSENG, G. C. & LEWIS, D. A. 2009. A molecular signature of depression in the amygdala. *Am J Psychiatry,* 166**,** 1011-24.

SINGHAL, G. & BAUNE, B. T. 2018. Chapter 8 - Do Chemokines Have a Role in the Pathophysiology of Depression? *In:* BAUNE, B. T. (ed.) *Inflammation and Immunity in Depression.* Academic Press.

SLATTERY, D. A., HUDSON, A. L. & NUTT, D. J. 2004. Invited review: the evolution of antidepressant mechanisms. *Fundamental & Clinical Pharmacology,* 18**,** 1-21.

SMITH, K. M., RENSHAW, P. F. & BILELLO, J. 2013. The diagnosis of depression: current and emerging methods. *Comprehensive psychiatry,* 54**,** 1-6.

SPIJKER, S., VAN ZANTEN, J. S., DE JONG, S., PENNINX, B. W., VAN DYCK, R., ZITMAN, F. G., SMIT, J. H., YLSTRA, B., SMIT, A. B. & HOOGENDIJK, W. J. 2010. Stimulated gene expression profiles as a blood marker of major depressive disorder. *Biol Psychiatry,* 68**,** 179-86.

STEPTOE, A., KUNZ-EBRECHT, S. R. & OWEN, N. 2003. Lack of association between depressive symptoms and markers of immune and vascular inflammation in middle-aged men and women. *Psychol Med,* 33**,** 667-74.

SULLIVAN, P. F., NEALE, M. C. & KENDLER, K. S. 2000. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry,* 157**,** 1552-62.

SZEBENI, A., SZEBENI, K., DIPERI, T., CHANDLEY, M. J., CRAWFORD, J. D., STOCKMEIER, C. A. & ORDWAY, G. A. 2014. Shortened telomere length in white matter oligodendrocytes in major depression: Potential role of oxidative stress. *International Journal of Neuropsychopharmacology,* 17**,** 1579-1589.

TAN, A. C. & GILBERT, D. 2003. Ensemble machine learning on gene expression data for cancer classification. *Appl Bioinformatics.*

TAYLOR, D., SPARSHATT, A., VARMA, S. & OLOFINJANA, O. 2014. Antidepressant efficacy of agomelatine: meta-analysis of published and unpublished studies. *BMJ : British Medical Journal,* 348.

TOCHIGI, M., IWAMOTO, K., BUNDO, M., SASAKI, T., KATO, N. & KATO, T. 2008. Gene expression profiling of major depression and suicide in the prefrontal cortex of postmortem brains. *Neurosci Res,* 60**,** 184-91.

TONELLI, L. H., STILLER, J., RUJESCU, D., GIEGLING, I., SCHNEIDER, B., MAURER, K., SCHNABEL, A., MÖLLER, H. J., CHEN, H. H. & POSTOLACHE, T. T. 2008.

Elevated cytokine expression in the orbitofrontal cortex of victims of suicide. *Acta Psychiatrica Scandinavica,* 117**,** 198-206.

TRULLAS, R. & SKOLNICK, P. 1990. Functional antagonists at the NMDA receptor complex exhibit antidepressant actions. *Eur J Pharmacol,* 185**,** 1-10.

UHER, R., PAYNE, J. L., PAVLOVA, B. & PERLIS, R. H. 2014. Major depressive disorder in DSM-5: implications for clinical practice and research of changes from DSM-IV. *Depress Anxiety,* 31**,** 459-71.

VAN DEN AKKER, E. B., PASSTOORS, W. M., JANSEN, R., VAN ZWET, E. W., GOEMAN, J. J., HULSMAN, M., EMILSSON, V., PEROLA, M., WILLEMSEN, G., PENNINX, B. W., HEIJMANS, B. T., MAIER, A. B., BOOMSMA, D. I., KOK, J. N., SLAGBOOM, P. E., REINDERS, M. J. & BEEKMAN, M. 2014. Meta-analysis on blood transcriptomic studies identifies consistently coexpressed protein-protein interaction modules as robust markers of human aging. *Aging Cell,* 13**,** 216-25.

VOINEAGU, I., WANG, X., JOHNSTON, P., LOWE, J. K., TIAN, Y., HORVATH, S., MILL, J., CANTOR, R., BLENCOWE, B. J. & GESCHWIND, D. H. 2011. Transcriptomic Analysis of Autistic Brain Reveals Convergent Molecular Pathology. *Nature,* 474**,** 380-384.

VREEBURG, S. A., HOOGENDIJK, W. J., VAN PELT, J., DERIJK, R. H., VERHAGEN, J. C., VAN DYCK, R., SMIT, J. H., ZITMAN, F. G. & PENNINX, B. W. 2009. Major depressive disorder and hypothalamic-pituitary-adrenal axis activity: results from a large cohort study. *Arch Gen Psychiatry,* 66**,** 617-26.

WALKER, E. M., JR. & WALKER, S. M. 2000. Effects of iron overload on the immune system. *Ann Clin Lab Sci,* 30**,** 354-65.

WANDERS, R. J. & WATERHAM, H. R. 2006. Biochemistry of mammalian peroxisomes revisited. *Annu Rev Biochem,* 75**,** 295-332.

WANG, J., HODES, G. E., ZHANG, H., ZHANG, S., ZHAO, W., GOLDEN, S. A., BI, W., MENARD, C., KANA, V., LEBOEUF, M., XIE, M., BREGMAN, D., PFAU, M. L., FLANIGAN, M. E., ESTEBAN-FERNÁNDEZ, A., YEMUL, S., SHARMA, A., HO, L., DIXON, R., MERAD, M., HAN, M.-H., RUSSO, S. J. & PASINETTI, G. M. 2018. Epigenetic modulation of inflammation and synaptic plasticity promotes resilience against stress in mice. *Nature Communications,* 9**,** 477.

WANG, L., LOCKSTONE, H. E., GUEST, P. C., LEVIN, Y., PALOTAS, A., PIETSCH, S., SCHWARZ, E., RAHMOUNE, H., HARRIS, L. W., MA, D. & BAHN, S. 2010. Expression profiling of fibroblasts identifies cell cycle abnormalities in schizophrenia. *J Proteome Res,* 9**,** 521-7.

WANG, T., JI, Y. L., YANG, Y. Y., XIONG, X. Y., WANG, I. M., SANDFORD, A. J., LIANG, Z. A. & HE, J. Q. 2015. Transcriptomic profiling of peripheral blood CD4(+) T-cells in asthmatics with and without depression. *Gene,* 565**,** 282-7.

WEBER, D. & O'BRIEN, K. 2017. Cancer and Cancer-Related Fatigue and the Interrelationships With Depression, Stress, and Inflammation. *Journal of Evidence-based Complementary & Alternative Medicine,* 22**,** 502-512.

WEI, Z., WANG, W., BRADFIELD, J., LI, J., CARDINALE, C., FRACKELTON, E., KIM, C., MENTCH, F., VAN STEEN, K., VISSCHER, PETER M., BALDASSANO, ROBERT N. & HAKONARSON, H. 2013. Large Sample Size, Wide Variant Spectrum, and Advanced Machine-Learning Technique Boost Risk Prediction for Inflammatory Bowel Disease. *The American Journal of Human Genetics,* 92**,** 1008-1012.

WEINBERG, E. 1996. Iron withholding: a defense against viral infections. *Biometals,* 9**,** 393-399.

WEISKOPF, D., WEINBERGER, B. & GRUBECK-LOEBENSTEIN, B. 2009. The aging of the immune system. *Transpl Int,* 22**,** 1041-50.

WHITEFORD, H. A., DEGENHARDT, L., REHM, J., BAXTER, A. J., FERRARI, A. J., ERSKINE, H. E., CHARLSON, F. J., NORMAN, R. E., FLAXMAN, A. D., JOHNS, N., BURSTEIN, R., MURRAY, C. J. & VOS, T. 2013. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet,* 382**,** 1575-86.

WHITEFORD, H. A., FERRARI, A. J., DEGENHARDT, L., FEIGIN, V. & VOS, T. 2015. The Global Burden of Mental, Neurological and Substance Use Disorders: An Analysis from the Global Burden of Disease Study 2010. *PLoS ONE,* 10**,** e0116820.

WISE, T., CLEARE, A. J., HERANE, A., YOUNG, A. H. & ARNONE, D. 2014. Diagnostic and therapeutic utility of neuroimaging in depression: an overview. *Neuropsychiatric Disease and Treatment,* 10**,** 1509-1522.

WIUM-ANDERSEN, M., ØRSTED, D., NIELSEN, S. & NORDESTGAARD, B. 2013. Elevated c-reactive protein levels, psychological distress, and depression in 73 131 individuals. *JAMA Psychiatry,* 70**,** 176-184.

WRAY, N. R., RIPKE, S., MATTHEISEN, M., TRZASKOWSKI, M., BYRNE, E. M., ABDELLAOUI, A., ADAMS, M. J., AGERBO, E., AIR, T. M., ANDLAUER, T. M. F., BACANU, S. A., BAEKVAD-HANSEN, M., BEEKMAN, A. F. T., BIGDELI, T. B., BINDER, E. B., BLACKWOOD, D. R. H., BRYOIS, J., BUTTENSCHON, H. N., BYBJERG-GRAUHOLM, J., CAI, N., CASTELAO, E., CHRISTENSEN, J. H., CLARKE, T. K., COLEMAN, J. I. R., COLODRO-CONDE, L., COUVY-DUCHESNE, B., CRADDOCK, N., CRAWFORD, G. E., CROWLEY, C. A., DASHTI, H. S., DAVIES, G., DEARY, I. J., DEGENHARDT, F., DERKS, E. M., DIREK, N., DOLAN, C. V., DUNN, E. C., ELEY, T. C., ERIKSSON, N., ESCOTT-PRICE, V., KIADEH, F. H. F., FINUCANE, H. K., FORSTNER, A. J., FRANK, J., GASPAR, H. A., GILL, M., GIUSTI-RODRIGUEZ, P., GOES, F. S., GORDON, S. D., GROVE, J., HALL, L. S., HANNON, E., HANSEN, C. S., HANSEN, T. F., HERMS, S., HICKIE, I. B., HOFFMANN, P., HOMUTH, G., HORN, C., HOTTENGA, J. J., HOUGAARD, D. M., HU, M., HYDE, C. L., ISING, M., JANSEN, R., JIN, F., JORGENSON, E., KNOWLES, J. A., KOHANE, I. S., KRAFT, J., KRETZSCHMAR, W. W., KROGH, J., KUTALIK, Z., LANE, J. M., LI, Y., LI, Y., LIND, P. A., LIU, X., LU, L., MACINTYRE, D. J., MACKINNON, D. F., MAIER, R. M., MAIER, W., MARCHINI, J., MBAREK, H., MCGRATH, P., MCGUFFIN, P., MEDLAND, S. E., MEHTA, D., MIDDELDORP, C. M., MIHAILOV, E., MILANESCHI, Y., MILANI, L., MILL, J., MONDIMORE, F. M., MONTGOMERY, G. W., MOSTAFAVI, S., MULLINS, N., NAUCK, M., NG, B., et al. 2018. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. 50**,** 668-681.

WRIGHT, F. A. & SULLIVAN, P. F. 2014. Heritability and genomics of gene expression in peripheral blood. 46**,** 430-7.

XIE, Y., WANG, X. & STORY, M. 2009. Statistical methods of background correction for Illumina BeadArray data. *Bioinformatics,* 25**,** 751-7.

XU, F., YANG, J., CHEN, J., WU, Q., GONG, W., ZHANG, J., SHAO, W., MU, J., YANG, D., YANG, Y., LI, Z. & XIE, P. 2015. Differential co-expression and regulation analyses reveal different mechanisms underlying major depressive disorder and subsyndromal symptomatic depression. *BMC Bioinformatics,* 16**,** 112.

YANG, Y., HAN, L., YUAN, Y., LI, J., HEI, N. & LIANG, H. 2014. Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nature communications,* 5**,** 3231-3231.

YESAVAGE, J. A., BRINK, T. L., ROSE, T. L., LUM, O., HUANG, V., ADEY, M. & LEIRER, V. O. 1982a. Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research,* 17**,** 37-49.

YESAVAGE, J. A., BRINK, T. L., ROSE, T. L., LUM, O., HUANG, V., ADEY, M. & LEIRER, V. O. 1982b. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res,* 17**,** 37-49.

YI, Z., LI, Z., YU, S., YUAN, C., HONG, W., WANG, Z., CUI, J., SHI, T. & FANG, Y. 2012. Blood-based gene expression profiles models for classification of subsyndromal symptomatic depression and major depressive disorder. *PLoS One,* 7**,** e31283.

ZANONI, I. & GRANUCCI, F. 2013. Role of CD14 in host protection against infections and in metabolism regulation. *Frontiers in cellular and infection microbiology,* 3**,** 32-32.

ZHANG, B. & HORVATH, S. 2005. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol,* 4**,** Article17.

ZHANG, M., ZHANG, L., ZOU, J., YAO, C., XIAO, H., LIU, Q., WANG, J., WANG, D., WANG, C. & GUO, Z. 2009. Evaluating reproducibility of differential expression discoveries in microarray studies by considering correlated molecular changes. *Bioinformatics,* 25**,** 1662-1668.

ZHENG, C.-H., YUAN, L., SHA, W. & SUN, Z.-L. 2014. Gene differential coexpression analysis based on biweight correlation and maximum clique. *BMC Bioinformatics,* 15**,** 1-7.

List of Figures

Figure 1.1. Proportions of DALYs explained by mental and substance use disorders globally in 2010 ... 9

Figure 1.2. Proportions of DALYs for mental and substance use disorders in Australia in 2015 ... 10

Figure 2.1. Selection process flowchart ... 27

Figure 2.2. Workflow chart ... 28

Figure 2.3. Replicated differentially expressed genes identified in periphery cells in depression ... 36

Figure 2.4. The 25 replicated differentially expressed genes identified genes identified in the brain are involved in depression and cardiovascular diseases ... 41

Figure 2.5. The 6 replicated differentially expressed genes identified genes identified in the periphery are involved in depression and cardiovascular diseases ... 42

Figure 2.6. Peroxisome and the role of *PXMP2* in lipid and reactive oxygen species (ROS) metabolism ... 50

Figure 3.1. Heatmap plot of correlations between the module eigengenes depression subtypes ... 67

Figure 3.2. Scatterplots depicting correlations between Module Membership (MM) and Gene Significance (GS) ... 68

Figure 4.1. Balancing training dataset using SMOTE ... 89

Figure 4.2. Module membership distribution plot visualizing the relative importance of the modules for recurrent MDD ... 90

Figure 4.3. The *TFRC* immune-related functional relationships ... 93

Figure 5.1. Differentially expressed genes in recurrent MDD ... 115

Figure 5.2. Interaction network for *CD14*... 119

Figure 5.3. A candidate pathway network ... 123

Figure 6.1. Summary of the findings on investigation of transcriptome signature of depression ... 126

List of Tables

Table 2.1. The 57 replicated differentially genes expressed mapped to the 6 cortical areas in depression 30

Table 3.1. Summary of MDD patient subgroups 65

Table 3.2. Pathway enrichment analysis on the modules associated with recurrent MDD in older adults 70

Table 3.3. Signaling Pathway Impact Analysis (SPIA) on the modules associated with recurrent MDD in older adults 71

Table 5.1. Studies on the dysregulation of immune genes and oxidative stress factors in brain tissues in depression 101

Table 5.2. Studies on the dysregulation of immune genes in peripheral tissues 108

Table 5.3. Thirteen KEGG pathways associated with recurrent MDD 112

Table 5.4. The top 6 differentially expressed genes in recurrent MDD 116