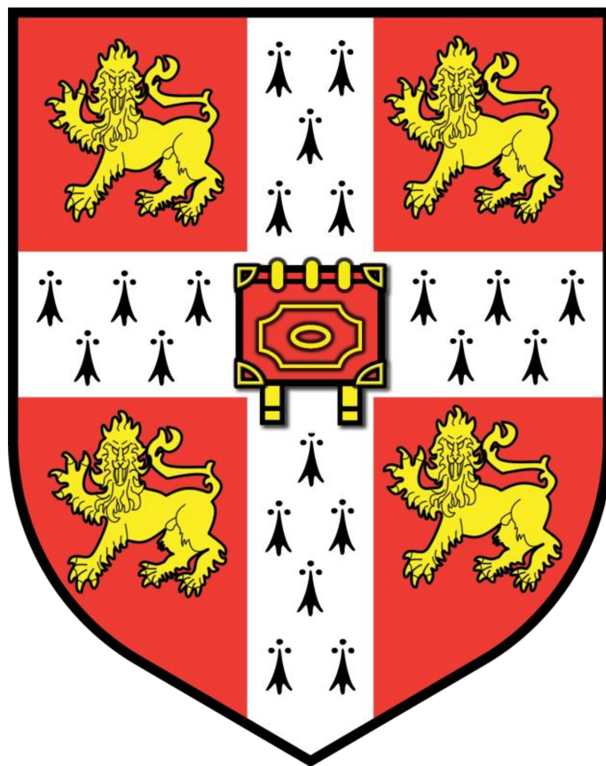
Central and peripheral molecular profiling of sex differences in schizophrenia, major depressive disorder, and controls



Jordan Marion Ramsey

Churchill College

May 2015

This dissertation is submitted for the degree of Doctor of Philosophy

To my family.

Acknowledgements

This thesis would not have been possible without the support of my supervisor, Prof Sabine Bahn. She provided me with guidance and interesting projects and funded me during the last couple months of my PhD. Her support has been crucial for my PhD work and career development. Most of the funding for my PhD was from the Edmonton Churchill Scholarship. In addition to their generous funding, the encouragement I received from my fellow Edmonton Churchill Scholars during my time here has been invaluable. I would also like to thank the Cambridge Commonwealth Trust for their support.

Dr Emanuel Schwarz gave me the first taste for working in this field and helped me through my first research projects on serum molecular sex differences and sex-dependent differences in schizophrenia. His creative take on this work inspired me to keep learning and coming up with new ideas. Dr Jason Cooper took me under his wing and my thesis would not be what it is without his insightful comments and discussions of all chapters in this thesis. I learned an immense amount in a short period with his help and he has helped to shape me as a scientist. He continues to support me and my career, give me confidence in my work, and continuously looks out for my best interests. I cannot thank him enough for that. I would also like to thank all other members of the Bahn lab (past and present).

This work would not have been possible without our collaborators: Nico JM van Beveren, F Markus Leweke, Matthias Rothermundt, Bernhard Bogerts, Johann Steiner, and Brenda Penninx. Microarray data used in this thesis was made available by the Stanley Medical Research Institute (SMRI), Harvard Brain Tissue Resource Centre, Jacqueline de Belleroche (Imperial College London, dataset GSE17612 deposited in the Gene Expression Omnibus (GEO)), and Elizabeth A Thomas (The Scripps Research Institute, dataset GSE21138 deposited in the GEO). I am very grateful to all patients and controls who participated in these studies.

Finally, I would like to give my heartfelt thanks to all my friends and family who supported me throughout my PhD and put up with me while I was writing up. Churchill College has been a home away from home during my time here. I have shared some truly lovely times with my housemates Ana, Hugo, Tim, and Matt and my Aussies Ryan, Jess, and Ash. My Mom and Dad have always been there for me in any way they can be, as have my brother, Hart and sister, Elizabeth. I would also like to thank my confidante Krista and partner Dave. I expect a lot from him and he always does his best to deliver, all the while keeping me a sane, balanced, and better person.

Declaration and Statement of Length

The work in this thesis is entirely my own, except where stated in the **Acknowledgements** or elsewhere in the text. This thesis, in its entirety or in part, has not been submitted for any other degree, diploma, or qualification at any university or other institution. This text does not exceed the 65,000 word limit and does not contain more than 150 figures, as required by the Degree Committee of Engineering.

Cambridge, United Kingdom, May 2015

Publications arising from this work

J. M. Ramsey, E. Schwarz, P.C. Guest, N.J.M. van Beveren, F.M. Leweke, M. Rothermundt, B. Bogerts, J. Steiner, L. Ruta, S. Baron-Cohen, S. Bahn, Molecular sex differences in human serum. *PLOS ONE*. **7**, e51504 (2012).

J. M. Ramsey, E. Schwarz, P.C. Guest, N.J.M. van Beveren, F.M. Leweke, M. Rothermundt, B. Bogerts, J. Steiner, S. Bahn, Distinct molecular phenotypes in male and female schizophrenia patients. *PLOS ONE*. **8**, e78729 (2013).

J. M. Ramsey, E. Schwarz, P.C. Guest, N.J.M. van Beveren, F.M. Leweke, M. Rothermundt, B. Bogerts, J. Steiner, S. Bahn, EPA-1234 – Distinct molecular phenotypes in male and female schizophrenia patients. *Eur. Psychiatry*. **29**, 1 (2014).

J. M. Ramsey, S. Bahn, P. 8. b. 023 Overlapping and distinct sex-specific serum molecular profiles in schizophrenia, depression, bipolar disorder, and Asperger syndrome. *Eur. Neuropsychopharmacol.* **24**, S746–S747 (2014).

Abstract

Schizophrenia and major depressive disorder (MDD) are disabling, poorly understood conditions with inadequate diagnostic and therapeutic technologies. They are characterized by clinical and epidemiological heterogeneity between male and female patients. Females have a higher prevalence of MDD, and in schizophrenia males have a higher incidence, earlier onset, more negative symptoms, and a worse prognosis and response to antipsychotic medication in schizophrenia. A better understanding of sex differences in these patients is needed to provide insight into biological mechanisms in these conditions and develop better diagnostic tools and therapies for males and females.

This thesis sought to examine molecular sex differences in control subjects and in schizophrenia and MDD patients by profiling serum and prefrontal cortex (PFC) brain tissue. Multiplex immunoassay and microarray technologies were used to measure concentrations of proteins and small molecules in serum and gene expression in brain, respectively. Extensive differences in serum molecular concentrations were found between males and females experiencing a menstrual cycle, using oral contraceptives, and after menopause. These were involved in immune, metabolic, and endocrine processes and may play a role in driving sex differences in susceptibility and other characteristics of mental disorders. Furthermore, the hypothalamic-pituitary-gonadal (HPG) axis was altered in MDD and first onset, antipsychotic naive schizophrenia patients. Biological processes involving response to steroid hormone stimulus and estradiol were enriched among differentially expressed genes in the PFC in schizophrenia. However, investigation of sex-dependent differences in PFC gene expression in schizophrenia was limited by the small number of female samples available. Finally, higher serum concentrations of a number of immune and inflammatory molecules were associated with MDD and schizophrenia in males only and this may have consequences for immune hypotheses of these disorders.

The research undertaken in this thesis has advanced the understanding of the mechanisms potentially involved in schizophrenia and MDD in males and females. Furthermore, sex and female hormonal status should routinely be taken into account during design and analysis of psychiatric molecular data to increase the reproducibility of such studies. This research shows that failing to consider these in biomarker studies can result in a high proportion of false positive findings and obscure important sex-dependent markers of mental disorders.

Contents

Chapter 1 Introduction ... 1
1.1 The burden and challenges of mental disorders ... 1
1.2 Sex ... 3
1.2.1 Endocrine and nervous systems ... 3
1.2.2 Other systems ... 7
1.3 Gender ... 9
1.4 Diagnosis of mental disorders ... 10
1.5 Research tools in mental disorders ... 11
1.6 Major depressive disorder (MDD) ... 12
1.6.1 Criteria for diagnosis ... 13
1.6.2 Course and treatment ... 14
1.6.3 Risk factors ... 14
1.6.4 Pathology ... 15
1.6.5 Hypotheses ... 16
1.7 Sex and gender differences in MDD ... 18
1.7.1 Hypotheses and pathology ... 18
1.8 Schizophrenia ... 21
1.8.1 Criteria for diagnosis ... 22
1.8.2 Course and treatment ... 23
1.8.3 Risk factors ... 24
1.8.4 Pathology ... 24
1.8.5 Hypotheses ... 25
1.9 Sex and gender differences in schizophrenia ... 29
1.9.1 Pathology ... 30
1.9.2 Hypotheses ... 31
1.10 Study aims and outline ... 32
Chapter 2 Methods ... 35
2.1. Human DiscoveryMAP® multiplex immunoassay ... 35
2.1.1 Advantages ... 37
2.1.2 Disadvantages ... 37
2.1.3 Clinical samples ... 37
2.1.4 Multiplex immunoassay data ... 37
2.2 Affymetrix GeneChip® DNA microarrays ... 38
2.2.1 Advantages ... 39

2.2.2 Disadvantages 40
2.2.3 Clinical samples 40
2.2.4 Microarray data 41
2.3 Statistics 42
2.3.1 ComBat 42
2.3.2 Mahalanobis distance 42
2.3.3 Principal component analysis (PCA) 43
2.3.4 Linear regression 43
2.3.5 Robust regression 45
2.3.6 Generalized least squares (GLS) 45
2.3.7 Logistic regression 46
2.3.8 Interactions 47
2.3.9 Model selection 49
2.3.10 False discovery rate (FDR) 49
2.3.11 Random forests 50
2.3.12 Classification performance 51
2.4 Gene Ontology (GO) terms 52
2.4.1 GO term enrichment 53
Chapter 3 Variations in serum molecular concentrations associated with sex and female hormonal status 54
3.1 Introduction 54
3.1.1 Aims and objectives 55
3.2 Methods 56
3.2.1 Clinical Samples 56
3.2.2 Data pre-processing 57
3.2.3 Sample exclusion 58
3.2.4 Data analysis 58
3.2.5 Simulation of group imbalance of sex and female hormonal status 62
3.2.6 Classification of sex and female hormonal status 62
3.3 Results 62
3.4 Discussion 74
Chapter 4 Sex differences in serum markers of major depressive disorder in the Netherlands Study of Depression and Anxiety 79
4.1 Introduction 79
4.1.1 Aims and objectives 80
4.2 Methods 81
4.2.1 Clinical Samples 81

4.2.2 Data pre-processing ... 84
4.2.3 Sample exclusion ... 84
4.2.4 Data analysis ... 85
4.2.5 Joint analyte effects and classification ... 86
4.2.6 Female HPG axis function ... 86
4.3 Results ... 86
4.4 Discussion ... 92
Chapter 5 Sex-dependent serum molecular differences in first onset, antipsychotic naive schizophrenia ... 96
5.1 Introduction ... 96
5.1.1 Aims and objectives ... 96
5.2 Materials and Methods ... 97
5.2.1 Clinical samples ... 97
5.2.2 Data pre-processing ... 98
5.2.3 Hormonal status ... 100
5.2.4 Data analysis ... 100
5.2.5 Classification of schizophrenia patients and controls ... 101
5.3 Results ... 102
5.4 Discussion ... 109
Chapter 6 A meta-analysis of gene expression profiling of the prefrontal cortex: schizophrenia, sex, and sex-schizophrenia diagnosis interaction effects ... 113
6.1 Introduction ... 113
6.1.1 Aims and objectives ... 115
6.2 Materials and Methods ... 115
6.2.1 Clinical samples ... 115
6.2.2 Data pre-processing ... 119
6.2.3 Data analysis ... 120
6.2.4 Potential confounders ... 121
6.2.5 Comparison with serum results ... 121
6.2.6 GO terms ... 121
6.3 Results ... 122
6.3.1 Differential PFC gene expression in schizophrenia ... 123
6.3.2 Sex-biased gene expression ... 135
6.3.3 Sex-schizophrenia diagnosis interactions in PFC gene expression ... 135
6.4 Discussion ... 146
6.4.1 Differential PFC gene expression in schizophrenia ... 146
6.4.2 Sex-biased gene expression ... 149

6.4.3 Sex-schizophrenia diagnosis interactions in PFC gene expression 150
Chapter 7 Final discussion 153
7.1 Summary of major findings 153
7.1.1 Inflammation and immune response 153
7.1.2 HPG axis 155
7.1.3 Implications for biomarker and microarray studies 156
7.2 Limitations and future outlook 157
7.3 Conclusions 161
Appendix 163
References 166

List of Figures

Figure 1.1. A unified model of sexual differentiation of mammalian tissues, adapted from *Arnold (2009).* 3
Figure 1.2. The female (left) and male (right) HPG axes, from *Hiller-Sturmhöfel and Bartke (1998).* 4
Figure 1.3. The female menstrual cycle. 5
Figure 1.4. The HPA axis, from *Hiller-Sturmhöfel and Bartke (1998).* 6
Figure 1.5. The role of 17β-estradiol in immune function, from *Fish (2008)* 8
Figure 1.6. A vision for schizophrenia; from *Insel (2010).* 12
Figure 1.7. An affective, biological, and cognitive model explaining gender differences in prevalence for MDD from *Hyde et al (2008).* 21
Figure 1.8. Neurodevelopmental model of schizophrenia; from *Insel (2010)* 28
Figure 1.9. Male and female age-at-onset distributions for schizophrenia in the WHO Determinants of Outcome 10-country Study. 30
Figure 2.1. Overview of Luminex xMAP technology used to quantify serum concentration of proteins and small molecules 36
Figure 2.2. Illustration of a single Affymetrix GeneChip® feature. 39
Figure 2.3. Summary of the steps performed during an Affymetrix GeneChip® DNA microarray experiment 40
Figure 2.4. Illustration of main effects and interactions 48
Figure 2.5. Simulated decision tree for classifying an observation as either disease or control based on $\log_2$-transformed analyte concentrations. 51
Figure 2.6. Examples of ROC curves to assess classification performance and their AUCs. 52
Figure 3.1. ComBat-normalized 6Ckine and insulin-like growth factor binding protein (IGFBP)-5 analyte concentrations.. 64
Figure 3.2. Serum analytes elevated significantly in (A) females and (B) males 67
Figure 3.3. Serum analytes varying significantly with female hormonal status. 70
Figure 3.4. Principal component analysis (PCA) plot of NESDA control samples 71
Figure 3.5. Plot of random forest variable importance for classifying sex and female hormonal status. 73
Figure 4.1. Sample exclusion for patient groups and controls after data pre-processing 84
Figure 4.2. Analytes with significant interactions between $\log_2$-transformed serum concentration and sex in MDD compared to controls 89

Figure 4.3. Analytes with overlapping significant sex-log_2 concentration interactions between MDD, CMA, and remitted MDD compared to controls 90
Figure 4.4. ROC curve illustrating classification of MDD for males and females using repeated ten-fold cross validation of BIC forward stepwise logistic regression 91
Figure 4.5. Serum concentrations of testosterone and FSH in controls and current and remitted female MDD 92
Figure 5.1. Analytes with significantly different log_2-transformed serum concentration between schizophrenia patients and controls. 104
Figure 5.2. Log_2-transformed analytes with significant interactions between schizophrenia diagnosis and sex. 106
Figure 5.3. ROC curve illustrating classification of schizophrenia for (A) males and (B) females using random forests. 107
Figure 6.1. Lateral surface of the human PFC 119
Figure 6.2. Array quality weights used for data analysis. 122
Figure 6.3. Probe sets with significantly different intensity between schizophrenia patients and controls in the PFC 131
Figure 6.4. Probe sets with significantly different intensity between male and female control PFC. 139
Figure 6.5. Probe sets with sex-schizophrenia diagnosis interactions with $p<0.01$. 145
Figure 7.1. Model of development of the PCOS phenotype. 156

List of Tables

Table 1.1. Summary of differences between males and females in mental disorders. 2
Table 1.2. DSM-IV criteria for diagnosis of MDD. 13
Table 1.3. DSM-IV criteria for diagnosis of schizophrenia. 22
Table 2.1. Results of performing m hypothesis tests: correct and incorrect rejections or failures to reject the null hypothesis. 50
Table 3.1. Demographic, lifestyle, and health characteristics for (A) NESDA controls (training set) and (B) NESDA test set subjects. 60
Table 3.2. Serum analytes elevated in (A) females and (B) males 65
Table 3.3. Serum analytes varying significantly with female hormonal status 68
Table 3.4. Average percentage of false discoveries in groups with imbalances in sex and female hormonal status. 72
Table 3.5. Test error for random forest classifiers. 73
Table 4.1. Female (A) and male (B) demographic, health, and lifestyle characteristics for MDD, CMA, remitted MDD, and controls. 82
Table 4.2. Analytes with significant interactions between $\log_2$-transformed serum concentration and sex in MDD compared to controls.. 88
Table 4.3. Analytes with overlapping significant sex-$\log_2$ concentration interactions between MDD and CMA (A) and between MDD and remitted MDD (B) compared to controls........... 90
Table 5.1. Patient and control characteristics by cohort. 99
Table 5.2. Treated schizophrenia cohort characteristics. 100
Table 5.3. Log_2-transformed analytes with significant interactions between schizophrenia diagnosis and sex. 105
Table 5.4. Sex differences in associations of serum molecular concentrations with positive, negative, and general PANSS score ratings 108
Table 6.1. Patient and control characteristics by cohort. 117
Table 6.2. Probe sets with significantly different intensity between schizophrenia patients and controls in the PFC. 124
Table 6.3. Results of analysis with and without the use of array weights and robust regression in the presence of regression outlier(s). 132
Table 6.4. Enriched GO BP terms for differentially expressed genes in schizophrenia 132
Table 6.5. Comparison of PFC gene expression findings with serum findings for first onset, antipsychotic naive schizophrenia patients compared to controls in Chapter 5................. 134
Table 6.6. Probe sets with significantly different intensity between male and female control PFC. 136

Table 6.7. Enriched GO BP terms for genes with sex-biased expression.......................... 140
Table 6.8. Probe sets with sex-schizophrenia diagnosis interactions with *p*<0.01. 140
Table 7.1. Over-represented gene ontology (GO) biological process (BP) terms among proteins investigated in (A) Chapters 3-4 (NESDA cohort) and (B) Chapter 5 (schizophrenia cohorts)... 160

Abbreviations

5-HT2A	Serotonin 2A
AAT	α1-antitrypsin
ACE	Angiotensin-converting enzyme
ACTH	Adrenocorticotropic hormone
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
ANOVA	Analysis of variance
ATC	Anatomical Therapeutic Chemical
AUC	Area under the ROC curve
BA	Brodmann area
BAI	Beck Anxiety Inventory
BBB	Blood-brain-barrier
BDNF	Brain-derived neurotrophic factor
BIC	Bayesian information criterion
BMI	Body mass index
BP	Biological process
CA 19-9	Cancer antigen 19-9
CAD	Coronary artery disease
cAMP	Cyclic adenosine monophosphate
cDNA	Complementary deoxyribonucleic acid
CEA	Carcinoembryonic antigen
CIDI	Composite International Diagnostic Interview
CLIA	Clinical Laboratory Improvement Amendments
CMA	Comorbid MDD and anxiety disorder(s)
COMT	Catechol-O-methyltransferase
cRNA	Complementary ribonucleic acid
CRP	C-reactive protein
CSF	Cerebrospinal fluid
DALY	Disability-adjusted life year
DLPFC	Dorsolateral prefrontal cortex
DNA	Deoxyribonucleic acid
DSM	Diagnostic and Statistical Manual of Mental Disorders
EBI	European Bioinformatics Institute
ECT	Electroconvulsive therapy
EGF	Epidermal growth factor
ELISA	Enzyme-linked immunosorbent assays
ER	Estrogen receptor
FAI	Free androgen index
FDR	False discovery rate
FGA	First-generation antipsychotic
fMRI	Functional magnetic resonance imaging
fRMA	Frozen robust multi-array average
FSH	Follicle-stimulating hormone
GABA	γ-aminobutyric acid
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GC	Glucocorticoid
GEO	Gene Expression Omnibus
GH	Growth hormone
GLS	Generalized least squares
GnRH	Gonadotropin-releasing hormone
GO	Gene ontology
GWAS	Genome-wide association study

HBTRC	Harvard Brain Tissue Resource Centre
HGF	Hepatocyte growth factor
HPA	Hypothalamic-pituitary-adrenal
HPG	Hypothalamic-pituitary-gonadal
HPT	Hypothalamic-pituitary-thyroid
ICAM-1	Intercellular adhesion molecule-1
ICD	International Statistical Classification of Diseases and Related Health Problems
ID	Intellectual disability
IDS	Inventory of Depressive Symptomatology
IGF	Insulin-like growth factor
IGF-1R	Insulin-like growth factor type 1 receptor
IGFBP	Insulin-like growth factor-binding protein
IgM	Immunoglobulin M
IL	Interleukin
LH	Luteinizing hormone
MAO	Monoamine oxidase
MAOI	Monoamine oxidase inhibiter
MAPK	Mitogen-activated protein kinase
MCP	Monocyte chemotactic protein
MDC	Macrophage derived chemokine
MDD	Major depressive disorder
MET	Metabolic equivalent
MIP	Macrophage inflammatory protein
MIP-1α	Macrophage inflammatory protein-1α
MLE	Maximum likelihood estimation
MM	Mismatch
MMP-3	Matrix metalloproteinase-3
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
N	Number
NESD	Netherlands Study of Depression and Anxiety
NIMH	National Institute of Mental Health
NMDA	*N*-Methyl-D-aspartic acid
NMDAR	*N*-Methyl-D-aspartic acid receptor
NSAID	Non-steroidal anti-inflammatory drug
NUSE	Normalized Unscaled Standard Error
OC	Oral contraceptive pill
OGK	Orthogonalized Gnanadesikan-Kettenring
OLS	Ordinary least squares
OOB	Out-of-bag
OR	Odds ratio
P	P-value
PAI-1	Plasminogen activator inhibitor 1
PANSS	Positive and Negative Syndrome Scale
PC	Principal component
PCA	Principal component analysis
PCOS	Polycystic ovary syndrome
PCP	Phencyclidine
PD	Personality disorder
PFC	Prefrontal cortex
PM	Perfect match
PMI	Post-mortem interval
Q	Q-value
RDoC	Research Domain Criteria

RLE	Relative Log Expression
RMA	Robust multi-array average
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
SAP	Serum amyloid P-component
SCF	Stem cell factor
SGA	Second-generation antipsychotic
SHBG	Sex hormone-binding globulin
SMRI	Stanley Medical Research Institute
SNP	Single-nucleotide polymorphism
SNRI	Serotonin and noradrenalin reuptake inhibitor
SSRI	Selective serotonin reuptake inhibitor
STARD	Standards for Reporting of Diagnostic Accuracy
TBG	Thyroxine binding globulin
TCA	Tricyclic antidepressant
TFF3	Trefoil factor 3
TNFR	Tumor necrosis factor receptor
uPAR	Urokinase-type plasminogen activator receptor
VEGF	Vascular endothelial growth factor
vWF	Von Willebrand factor
WHO	World Health Organization
WPA	World Psychiatric Association

Remaining analyte abbreviations can be found in the **Appendix** (**Table A.1**).

Chapter 1 Introduction

1.1 The burden and challenges of mental disorders

A mental disorder is ‘a syndrome characterized by clinically significant disturbance in an individual’s cognition, emotion regulation, or behaviour that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning’, according to the Diagnostic and Statistical Manual of Mental Disorders (DSM)-5 (*1*). They include schizophrenia spectrum and psychotic disorders, bipolar disorders, depressive disorders (*e.g.,* major depressive disorder), anxiety disorders, neurodevelopmental disorders (*e.g.,* autism spectrum disorders), and a number of others. Heterogeneous in nature, they are characterized by a spectrum of distinct and often overlapping symptoms. These disabling and stigmatized conditions have a global lifetime prevalence of 18-36%, are associated with largely unsatisfactory diagnostic and therapeutic technologies (*2*), and are expected to cause an estimated cumulative global loss in economic output of US$ 16.3 trillion between 2011 and 2030 (*3*). The biological bases for these disorders have not yet been elucidated, though genetic and environmental, neurotrophic, neurochemical, endocrine, and infectious and immune factors have been implicated. However, a number of challenges hinder progress, including: the subjective basis for the diagnosis of mental disorders; use of different diagnostic manuals or other criteria frequently revised over time; inaccessibility of the brain; confounding with psychiatric medication use; high comorbidity with other mental disorders and physical illnesses; misdiagnosis and changes in diagnosis over time; ethical considerations; lack of validated animal models; and heterogeneous clinical presentation, epidemiology, and biological findings.

An important component of heterogeneity within mental disorders are sex and gender differences (defined in **Sections 1.2** and **1.3**) in prevalence and incidence, age of onset, symptoms, course, treatment response, and other clinical characteristics. A summary of differences between males and females in a subset of the major mental disorders can be found in **Table 1.1**. The importance of studying sex and gender differences in health is increasingly recognized in policy and health programming, and in the social and biomedical sciences (*4–6*). Initiatives to include them in research have taken place in Canada, the United States, and the European Union. Novel insights have been gained into disease mechanisms and treatment by considering sex and gender in research of neurological, cardiovascular, pulmonary, and autoimmune diseases that show prominent differences between men and women (*7*, *8*). Exploration of sex-based heterogeneity in the biological

underpinnings of mental disorders could yield novel insights into aetiology, pathophysiology, and treatment options, and provide new personalized medicine approaches. The limited research done in this field has found estrogen to be a promising treatment for female schizophrenia patients (*9*) and has discovered sex-specific biomarker profiles for Asperger syndrome (*10*). This thesis will focus on sex-dependent molecular differences in the serum and brain tissue of patients with schizophrenia and major depressive disorder (MDD), mental disorders with well-characterized and profound sex and gender differences. A more detailed discussion of these disorders and sex and gender differences in them is given, following a short introduction to biological and socio-cultural differences between men and women.

Table 1.1. Summary of differences between males and females in mental disorders.

Mental disorder	**Sex and gender differences**
Autism spectrum disorders *(Fombonne, 2005)* (*11*)	• Higher prevalence in males (~4:1) • Lower cognitive abilities in females
Schizophrenia *(Abel et al, 2010)* (*12*)	• Higher incidence in males (~1.4:1) • Later female age of onset • Better premorbid function, social course, prognosis, and response to antipsychotic medication in females • More affective symptoms in females; more negative symptoms in males
Major depressive and anxiety disorders *(Bromet et al, 2011* (*13*)*; Hawton, 2000* (*14*))	• More prevalent in females (~2:1) • Higher suicide rate in depressed males
Posttraumatic stress disorder *(Tolin and Foa, 2006)* (*15*)	• Higher risk to females following traumatic experience, especially after assault • Differences in vulnerability to types of trauma
Obsessive-compulsive disorder *(de Mathis et al, 2011)* (*16*)	• Earlier onset in males • More chronic course of illness in males, greater social impairment • Differences in symptoms and psychiatric comorbidities
Bipolar disorder *(Kawa et al, 2005)* (*17*)	• Later female age of onset • Higher prevalence of bipolar disorder II, rapid cycling, and mixed mania in females
Anorexia and bulimia nervosa *(Hudson et al, 2007)* (*18*)	• Higher prevalence in females (~3:1)
Substance use disorders *(Becker and Hu, 2007)* (*19*)	• Higher prevalence in males (~2:1) • Differences in drug abuse initiation, escalation of use, addiction, and relapse
Personality disorders (PDs) *(Cale and Lilienfeld, 2002* (*20*)*; Sansone and Sansone, 2011* (*21*)*; DSMs IV/5* (*1, 22*)*; Corbitt and Widiger (2006)* (*23*))	• Higher prevalence of antisocial PD in males (~3:1) • Schizoid, schizotypal, paranoid, narcissistic, obsessive-compulsive PDs may be more prevalent in males • Borderline personality, histrionic, dependent PDs may be more prevalent in females • Differences in personality traits, psychiatric comorbidity, and treatment of borderline PD

1.2 Sex

Sex refers to the biological and physiological attributes that characterize males and females. Sex differences in humans are present throughout life and range from obvious dichotomous characteristics to more subtle and less observable sexual dimorphisms. These emerge either as a direct result of sex chromosome effects or from indirect effects through levels of sex hormones, which influence growth, development, maintenance, and function of body systems (*8*, *24*). A summary of factors influencing biological expression of sex differences can be seen in **Figure 1.1**. Sex differences in a number of body systems and cellular processes relevant to mental disorders are described briefly in the following sections. These biological differences interact with socio-cultural factors to produce complex differences between men and women (*8*).

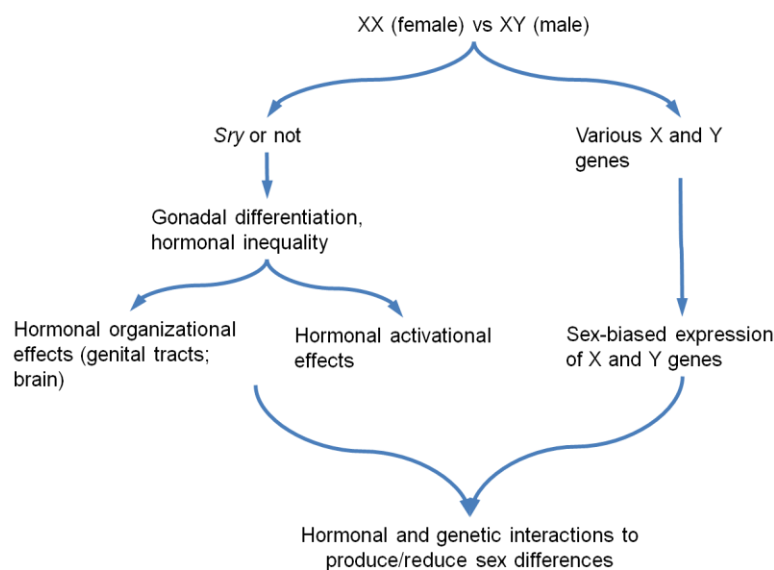


Figure 1.1. A unified model of sexual differentiation of mammalian tissues, adapted from *Arnold (2009)* (*25*)**.** *Reprinted with permission.*

1.2.1 Endocrine and nervous systems

Glands comprising the endocrine system secrete signalling molecules called hormones into the circulatory system, which regulate physiology and behaviour. Reproductive functions are controlled by complex feedback loops between the hypothalamus, pituitary, and gonads, referred to as the hypothalamic-pituitary-gonadal (HPG) axis and illustrated in **Figure 1.2.** Gonadotropin-releasing hormone (GnRH) from the hypothalamus stimulates production of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) in the anterior pituitary. Their release into the bloodstream influences male and female reproductive function, including production of sex steroids. Sex steroids include androgens, estrogens, and

progestogens, which are secreted from the gonads or adrenal glands or are converted from other sex steroids.

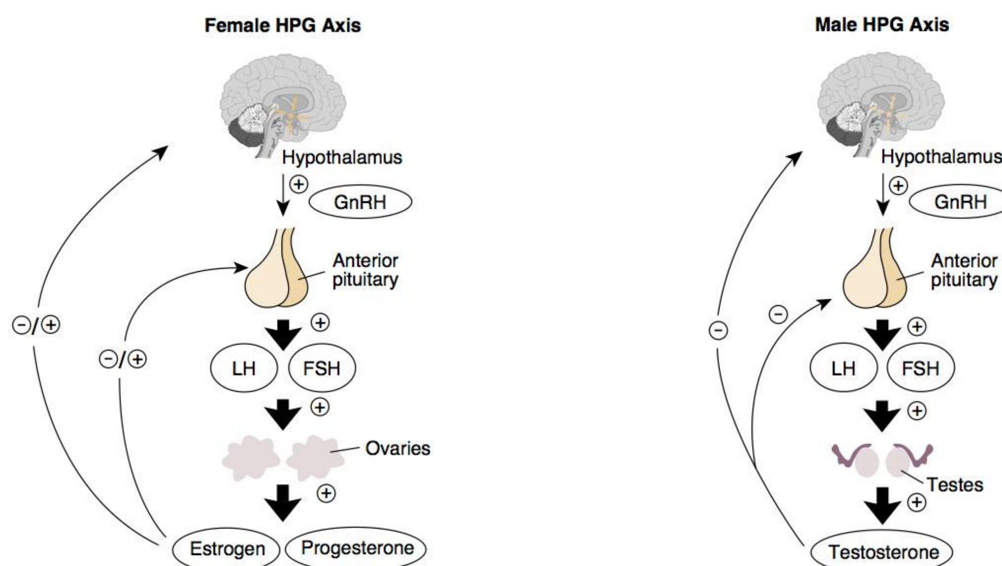


Figure 1.2. **The female (left) and male (right) HPG axes, from *Hiller-Sturmhöfel and Bartke (1998)*** (*26*). Abbreviations: HPG (hypothalamic-pituitary-gonadal); GnRH (gonadotropin-releasing hormone); LH (luteinizing hormone); FSH (follicle-stimulating hormone). *Reprinted with permission.*

The effects of sex steroids can generally be classified as organizational or activational (*24*) and occur through genomic and non-genomic mechanisms. Non-genomic effects cause rapid activation or repression of cell function when sex steroids modulate regulatory proteins on the cell membrane or in the cytoplasm (*27*). Genomic effects, on the other hand, affect medium- and long-term function and programming of cells and regulate gene expression and protein synthesis (*27*). In the male foetus, the *Sry* gene on the Y chromosome causes testes to grow and secretion of testosterone from them permanently affects brain and genital development. A female foetus develops without *Sry*, and a female brain and genitals develop. These permanent, differentiating effects constitute the organizational effects of sex steroids. Activational effects of sex hormones can be reversed and occur at puberty (together with further organizational effects) (*24*), during which time GnRH stimulates production and release of LH and FSH to in turn produce sex steroids. Sex steroids cause the genitals and other secondary sexual characteristics to develop. At this time, males begin to produce sperm and females begin menstruation.

Hormonal variations occur within the menstrual cycle and across the female lifespan, affecting many aspects of brain and body function. In the female menstrual cycle, as shown in **Figure 1.3,** an ovary releases an egg and the uterus is prepared for pregnancy in a

process that repeats approximately every 28 days (*28*). During the follicular phase (approximately days 1-14), ovarian follicles develop with elevated FSH. An estrogen-stimulated surge in LH, which occurs around day 12 and lasts 36-48 hours, causes the egg to be released from the follicle (called ovulation). In the luteal phase (approximately days 15-28), the follicle is transformed to a corpus luteum, which secretes progesterone and degenerates in the absence of pregnancy. Changes in body temperature and the uterine endometrium occur during the menstrual cycle. Hormonal contraception is used by a substantial number of reproductive-aged females in the developed world and can prevent ovulation, thicken cervical mucus, and may thin the endometrium. The most popular of these is the combined oral contraceptive pill (OC), containing estrogen (mostly ethinyl estradiol) and progestin. Pregnancy is marked by increases in estrogen and progesterone, which drop postpartum. Menopause marks the end of menstruation and generally occurs around age 51. Estrogen and progesterone ovarian production largely stops, while FSH and LH levels remain high.

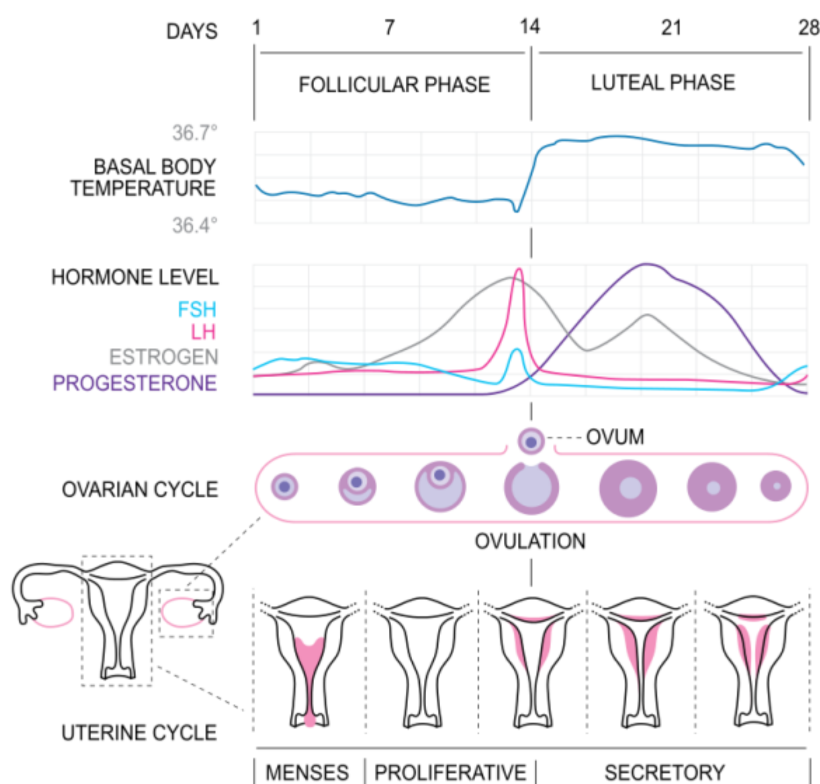


Figure 1.3. The female menstrual cycle. *Image courtesy of Isometrik; reprinted with permission.* **Abbreviations:** LH (luteinizing hormone); FSH (follicle-stimulating hormone).

In addition to their role in reproduction, sex hormones are important in a number of other physiological processes. The HPG axis interacts with other hormone systems, including the hypothalamic-pituitary-adrenal (HPA) axis (*29*). The HPA axis plays a key role in maintaining

homeostasis in the presence of physical and psychological stress and is illustrated in **Figure 1.4** (*30*). Altered activation of this network has been implicated in schizophrenia, depression and anxiety disorders, bipolar disorder, autism, and others (*31–34*). Sex differences in HPA axis development and functional cross-talk with the HPG axis may result in sex-dependent reactions to stress in systems regulated by these axes, including reproductive, nervous, musculoskeletal, cardiovascular, and immune systems, in addition to metabolism and electrolyte and water balance (*29*, *35–37*).

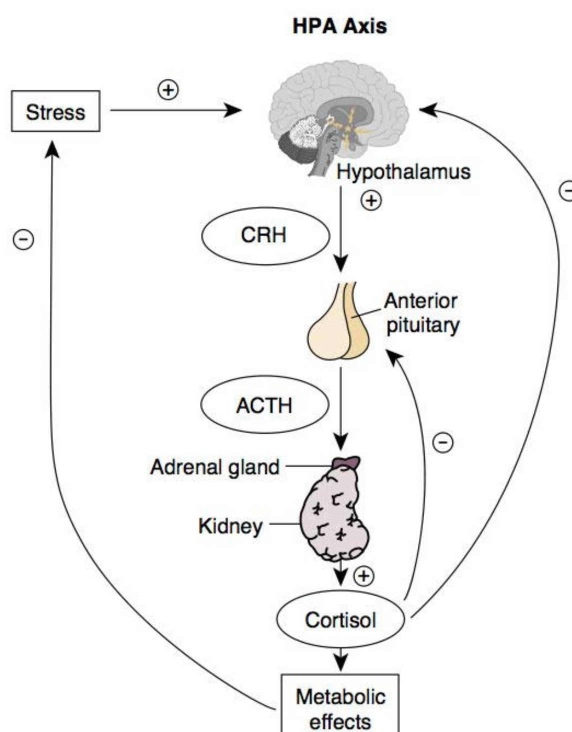


Figure 1.4. The HPA axis, from *Hiller-Sturmhöfel and Bartke (1998)* (*26*). Abbreviations: HPA (hypothalamic-pituitary-adrenal); CRH (corticotropin-releasing hormone); ACTH (adrenocorticotropic hormone). *Reprinted with permission.*

The influence of sex hormones on brain and behaviour is well documented. In addition to sex differences in mental disorders and neurological conditions, sex dimorphisms occur in memory, cognition, emotion, vision, hearing, face processing, pain perception, and other behaviours (*38*). Perinatal secretion of testosterone in males causes the brain to develop differently from females, and sex hormone receptors distributed throughout the brain influence neural function over the lifespan of an individual (*39*). Sex differences in humans and other species have been found in the volumes of certain brain areas, the number, architecture and activity of cells, neurochemical make-up, stress hormone action, and synaptic connectivity (*38*, *39*). Neurotransmitter systems like γ-aminobutyric acid (GABA), glutamate, acetylcholine, opioids, vasopressin, and monoamines differ between males and females (*38*, *40*, *41*). For instance, serotonin synthesis is more than 50% higher in human

males (*42*) and estradiol has antipsychotic effects in rats treated with amphetamine, which increases dopaminergic activity (*43*). In the hippocampus, vital for memory formation and recollection, glucocorticoid receptor affinity is reduced in females and sex hormones influence excitability of cells, dendritic structure, *N*-Methyl-D-aspartic acid (NMDA) receptor binding, and memory and learning processes (*38*). Studies in rats and monkeys have found chronic stress to cause more hippocampal damage to males (*38*) and estrogen has been found to be neuroprotective *in vivo*, mediated by its ability to reduce oxidative stress and neurotoxicity (*27*). The prefrontal cortex is involved in cognitive control and has a high concentration of sex hormone receptors and differences between males and females have been observed in brain activation during working memory tasks (*38*). Though much of the research on sex differences in neuroscience has focused on sex hormones, increasingly effects of the sex chromosomes are being investigated (*44*). X-chromosome genes influence development of intelligence, social cognition, and emotional regulation (*45*). X-linked intellectual disability (ID) is thought to account for 5-10% of ID in males (*46*).

1.2.2 Other systems

Sex differences exist in the functioning of other body systems and processes relevant to mental illness. Studies have suggested that immune and cardiovascular systems and metabolic processes are involved in mental disorders, often in a bi-directional manner (*47–52*).

1.2.2.1 Immune response

Females generally have a more robust immune response, which may be responsible for the higher prevalence of a number of autoimmune diseases and lower risk for many types of infection in females (*53*). The hormones progesterone, testosterone, and estrogens account for many of the observed sex differences in immune response. Though androgens largely suppress immune function, the role of estrogens in inflammation is complex and hormonal fluctuations in females can influence production of cytokines and chemokines, chemotaxis of cells to sites of inflammation, and development and function of immune cells (*53*), as illustrated in **Figure 1.5**. Females have more CD4+ T cells and estrogen increases the regulatory T-cell population, survival of autoreactive B-cells, and increase differentiation of dendritic cells from bone marrow (*53*). On the other hand, estrogen reduces the cytotoxicity of natural killer cells and has anti-inflammatory effects on monocytes, macrophages, and neutrophils (*53*). 17β-estradiol has been found to mediate anti-inflammatory pathways in microglial cells, which are brain macrophages that are necessary for early sexual differentiation and development (*54*). Estradiol reduces their activation, superoxide release, phagocytic activity, and recruitment of peripheral monocytes (*55*, *56*). Furthermore, sex

differences have been found in the number of microglia in developing brain regions, their morphology, and in their *ex vivo* inflammatory signalling responses in rodents (*54*).

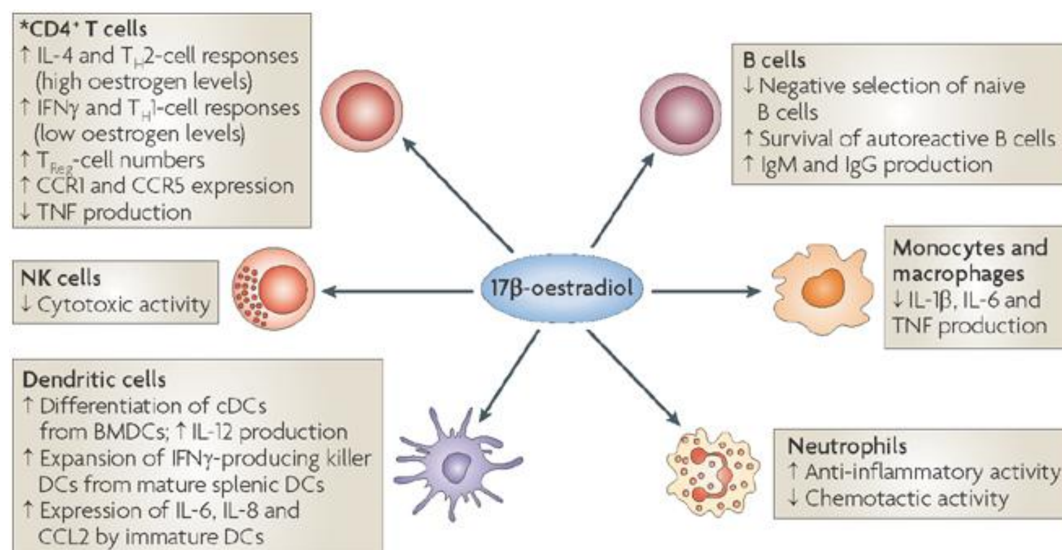


Figure 1.5. The role of 17β-estradiol in immune function, from *Fish (2008)* (*53*)**.** Abbreviations: IL (interleukin); IFN (interferon); Reg (regulatory); CCR (CC-chemokine receptor); TNF (tumour-necrosis factor); BMDC (bone-marrow-derived DC precursor); cDC (conventional dendritic cell); DC (dendritic cell); CCL (CC-chemokine ligand). *Reprinted with permission.*

1.2.2.2 Metabolism

Sex differences have been reported in glucose, lipid, and drug metabolism. Males are more sensitive to central insulin while females are more sensitive to leptin that control glucose metabolism and energy balance (*57*). Insulin-like growth factor type 1 receptor (IGF-1R) regulates blood glucose regulation, oxidative stress, and longevity in a sex dimorphic manner, with heterozygous knockout *Igf1r*$^{+/-}$ female mice showing increased lifespan and stress resistance and male mice showing a stronger glucose response than controls (*58*). Female rats are protected against mitochondrial oxidative stress, with lower oxidative damage and higher gene expression of antioxidant enzymes compared to males (*59*). Lipid metabolism differs greatly between the sexes, with human females generally having a higher percentage of body fat, greater oxidation of fat during exercise, and different fat mobilization and body fat distribution compared to males (*60*). While females have a greater tendency to deposit fat subcutaneously and in the gluteal-femoral region, males are more likely to build visceral fat, associated with increased risks for hypertension, cardiovascular disease, type 2 diabetes, and other disorders (*57*). Gonadal hormones play key roles in these sex differences, with testosterone increasing lipolysis and estradiol enhancing production of preadipocytes (*57*). Hormonal changes in menopause have been associated with increased

deposition of visceral fat and use of OCs have been shown to cause increased levels of serum triglyceride, low-density lipoprotein cholesterol, and insulin (*57*, *61*). Adipose tissue itself is also a source of leptin, adiponectin, resistin, cortisol, estradiol, testosterone, and inflammatory cytokines (*57*). In the brain, sex differences in glucose metabolism have been found to vary with brain region, with males having higher metabolism in temporal-limbic regions and the cerebellum and lower metabolism in cingulate regions (*62*). Estrogen and progesterone are believed to be protective and beneficial in recovery after brain injury and stroke, in part due to their capacity to reduce lipid peroxidation (*63*). Substantial sex differences in drug pharmacokinetics and pharmacodynamics have also been found and can be influenced by pregnancy, hormonal contraceptive use, menopause, and menstrual cycle (*64*). In particular, differences in hepatic enzymes between males and females are thought to be of great importance to pharmacokinetic differences (*64*).

1.2.2.3 Cardiovascular function

Estrogen is reported to have favourable effects on vascular function, with a better prognosis for most cardiovascular diseases and reduced incidence of atherosclerosis, coronary artery disease (CAD), and ischemic stroke in premenopausal women compared to men (*48*, *65*). However, estrogen replacement has also been shown to increase the risk of coronary heart disease in postmenopausal women, and combined OCs are associated with increased risk for venous thrombosis, myocardial infarctions, and ischemic strokes (*66–68*). In periphery and brain, estrogen affects vascular reactivity, promoting vasodilation and thereby increasing blood flow (*65*, *69*). Testosterone, on the other hand, promotes vasoconstriction (*65*, *70*). Consequently, blood pressure is higher in men than in premenopausal women, and increases after menopause (*69*). Estrogen has also been found to reduce oxidative stress and blood-brain barrier permeability, increase neurogenesis and angiogenesis, and inhibit formation of atherosclerotic plaques (*65*, *69*, *71–73*).

1.3 Gender

Gender refers to the socially constructed roles filled by men and women. While these roles may differ between cultures, biological attributes do not. A number of gender-based factors influence the differences in mental disorders between men and women. Gender inequalities in social status are highly prevalent, with larger gaps in developing countries. In general, women have lower levels of economic participation and opportunity, educational attainment, health and survival, and political empowerment, as quantified by the Global Gender Gap Index (*74*). However, men are more likely to engage in novel or intense activities (*75*) and

are less likely to seek help for health problems (*76*). Gender differences are present in vulnerabilities to different types of traumatic events (*77*), and perception of stress and coping style (*78*). The magnitude of gender differences in personality traits has been reported to vary across cultures and is greater in European and American cultures. Women were generally higher in neuroticism, agreeableness, warmth, and openness to feelings using the Revised NEO Personality Inventory, while men were higher in assertiveness and openness to new ideas (*79*). Social structural and behavioural determinants of health are also often different in men and women, adding to the complexity of gender differences in disease and disorders (*80*). Gender and sex are intricately tied and are both important in the emergence of differences between men and women in mental disorders.

1.4 Diagnosis of mental disorders

Mental disorders are currently diagnosed by interview carried out by a psychiatrist or other health or mental health professional. The majority of psychiatrists (approximately 80%) often or always use a formal classification system for diagnosis of patients (*81*). A 2010/2011 World Psychiatric Association (WPA)-World Health Organization (WHO) global survey of psychiatrists in 44 countries reported that Chapter V of the International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10; produced by the WHO) was most widely used (70%), followed by the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV; produced by the American Psychiatric Association; 23%), with the remainder using another classification system (*81*). The DSM has recently been revised and the 5th edition was released in May 2013 (*1*). A number of sources of variation can often reduce diagnostic reliability, and various methods have therefore been developed for clinical research and post-mortem studies. The Composite International Diagnostic Interview (CIDI) assesses mental disorders according to ICD and DSM criteria in a highly standardized way across various settings and cultures (*82*). The CIDI was intended for epidemiological studies, but is also used widely in the clinic and in research (*82*). Nevertheless, the issue of classification of mental disorders is contentious, with some arguing a dimensional approach is preferable to a categorical one (*83*). Evidence from clinical neuroscience and genetics suggests that diagnostic categories in the DSM and ICD may not align with pathophysiological mechanisms in mental disorders (*84*). The Research Domain Criteria (RDoC) project launched by the US National Institute of Mental Health (NIMH) conceptualizes mental disorders as brain disorders and aims to incorporate pathophysiological data into future classification schemes to improve patient outcomes (*84*). It defines five domains representing functional dimensions of behaviour and several units of

analysis, including genes, circuit, behaviour, and other measurements in psychopathology research (*85*). *White et al (2012)* (*86*) also argue that current scientific evidence supports the reclassification of mental and neurological illnesses as disorders of the central nervous system in diagnostic manuals.

1.5 Research tools in mental disorders

A number of tools and technologies are used to study biological psychiatry. Post-mortem brain studies are used to detect abnormalities in function and structure, including brain volume and cell size, density, and distribution. Structural neuroimaging is used to study structure of the brain in living patients and controls, while functional neuroimaging is used to investigate aspects of brain function (*87*). Brain activity can be measured using electrical currents, magnetic fields, regional cerebral blood flow and metabolism, and blood oxygenation. Neurotransmitter activity can also be studied using functional neuroimaging.

Genomic and transcriptomic studies can reveal risk genes and altered gene expression important in the aetiology and pathophysiology of mental disorders. Study of the functional products of genes (proteins) provides direct insight into the cellular mechanisms underlying mental disorders. Elucidating abnormalities in protein expression, structure, and function (proteomics) is needed to establish common pathways to the behavioural and psychological dysfunction in mental disorders caused by genetic vulnerability and environmental factors (*88*). Metabolic studies (metabolomics) can quantify perturbations in the cellular products of biochemical reactions that may underlie dysfunction in mental disorders. Central and peripheral omics studies have revealed a number of abnormalities in mental disorders (*89*). Peripheral studies allow easier and less invasive access to tissue from living patients and can mirror aspects of brain function.

In addition to direct patient measurements, cellular and animal models of mental disorders have been investigated. Fibroblasts from patients can be reprogrammed into human induced pluripotent stem cells and differentiated into neurons to study aspects of their function (*90*). Animal models of mental disorders can be used to study the physiological underpinnings of behaviours caused by specific environmental and biological manipulations.

A better understanding of mental disorders is imperative to improving patient diagnosis, therapy, and outcome. A clinical need has been identified for tools to prevent misdiagnosis, aid early diagnosis, and predict response to treatment (*91*). These may be achieved by a number of technologies and their translation into clinical practice, as seen in **Figure 1.6.** Increasingly, a need to address the heterogeneity in mental disorders is recognized, with an

emphasis on providing personalized care. The vision for schizophrenia in **Figure 1.6** was put forth by the director of the NIMH (*92*), but is also relevant for other mental disorders.

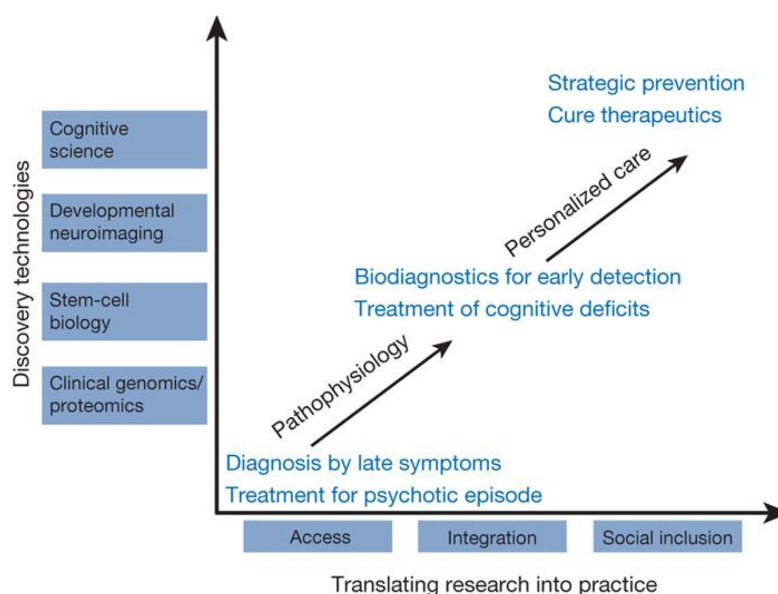


Figure 1.6. A vision for schizophrenia; from *Insel (2010)* (*92*)**.** *Reprinted with permission.*

1.6 Major depressive disorder (MDD)

Major depressive disorder (MDD) is a mental disorder characterized by heterogeneous symptoms including chronic low mood, loss of pleasure or interest in activities, and other physical and psychological symptoms. MDD is a highly prevalent and disabling condition with poor recognition in primary care settings (*93*, *94*). The average lifetime prevalence of MDD was 14.6% in high-income countries and 11.1% in middle- and low-income countries in a WHO World Mental Health survey published in 2011, with considerable variation between countries (*13*). The WHO predicts that by 2030 MDD will be the leading cause of disease burden worldwide, accounting for 6.2% of all disability-adjusted life years (DALYs), which measure years of healthy life lost due to disability and premature mortality (*95*). In Europe this has already occurred and annual costs of MDD have been estimated at €92 billion (59% direct costs, 41% indirect costs), accounting for approximately 12% of the cost of all brain disorders (*96*).

Other mental disorders share features of MDD. Dysthymia is a milder, chronic form of MDD now consolidated with chronic MDD as persistent depressive disorder in the DSM-5 (*1*). Bipolar disorder is characterized by episodes of depressed and elevated mood (mania or hypomania). Negative symptoms of schizophrenia may resemble depressive symptoms, while psychotic features may be present in severe MDD. MDD is frequently accompanied by

anxiety disorders, characterized by excess fear, anxiety, and associated behaviours (*1*). Anxiety disorders include social and generalized anxiety disorders, panic disorder, and phobias, such as agoraphobia and specific phobias. They are highly comorbid with each other and have different physical and psychological symptoms. MDD is also often comorbid with alcohol use disorders, eating disorders, borderline personality disorder, schizophrenia, and physical illnesses such as cardiovascular disease and type 2 diabetes (*1*, *48*, *52*, *97*, *98*). A relative risk of incidence of 1.5 for coronary artery disease is conferred independently by depression (*48*, *50*).

1.6.1 Criteria for diagnosis

The DSM-IV criteria for diagnosis of MDD (*22*) are presented in **Table 1.2**. The work in this thesis and most other published research has not been carried out with the most recent 5th edition.

Table 1.2. DSM-IV criteria for diagnosis of MDD.

Criteria
A. At least 5 of the following, during the same 2-week period, representing a change from previous functioning: at least one symptom is either (a) or (b) a. Depressed mood most of the day b. Diminished interest or pleasure in all, or almost all, activities c. Significant weight loss or gain d. Insomnia or hypersomnia e. Psychomotor agitation or retardation f. Fatigue or loss of energy g. Feelings of worthlessness or guilt h. Diminished ability to think or concentrate, or indecisiveness i. Recurrent thoughts of death, suicidal ideation, suicide attempt, or a specific plan for committing suicide
B. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
C. The episode is not attributable to the physiological effects of a substance or to another medical condition.
D. The occurrence of the major depressive episode is not better explained by schizoaffective disorder, schizophrenia, schizophreniform disorder, delusional disorder, or other specified and unspecified schizophrenia spectrum and other psychotic disorders.
E. There has never been manic episode or a hypomanic episode.
F. Symptoms are not better accounted for by bereavement.

Single or recurrent episodes are diagnosed with additional specifiers including severity, remission, chronic status, psychotic features (mood-congruent or mood-incongruent), atypical features, melancholic features, catatonia, postpartum onset, and seasonal pattern.

Changes in the DSM-5 include removal of the bereavement exclusion for a diagnosis of MDD and addition of specifiers, “with anxious distress” and “with mixed features”.

Premenstrual dysphoric disorder and disruptive mood disorder were also added in the DSM-5 and chronic MDD and dysthymia were combined as persistent depressive disorder, as stated above (*1*).

1.6.2 Course and treatment

The risk of MDD increases substantially with puberty, with peak incidence occurring in the 20s and 30s. Length of episode and course of illness varies considerably. In 40% of patients, recovery starts within three months of onset, and within one year 80% of patients will have begun to recover (*1*). Comorbid anxiety disorders are present in more than half of those suffering from depression, typically preceding development of MDD and associated with poorer psychiatric outcomes (*99*, *100*). Other factors associated with lower rates of recovery include psychotic features, personality disorders, substance use disorders, severe symptoms, and length of current episode (*1*). MDD is a highly recurrent condition with the risk of another episode at least 50% for individuals who have had one and 80% for those who have had two prior episodes (*101*). Certain patients may infrequently or never experience remission, while others may experience partial or complete recovery between episodes. MDD patients have high rates of attempted and completed suicides (*102*, *103*).

Pharmacological and other interventions are used in the treatment of MDD. The first generation of antidepressant medications were monoamine oxidase inhibiters (MAOIs) and tricyclic antidepressants (TCAs) that had several side effects. Selective serotonin reuptake inhibitor (SSRI) and serotonin and noradrenalin reuptake inhibitor (SNRI) antidepressant medications were developed later and have fewer adverse effects and better safety profiles. However, the effectiveness of antidepressant medication does not differ from placebo in cases of mild and moderate depression (*104*, *105*) and several weeks of treatment are required for therapeutic benefit to be seen. Psychological and psychosocial interventions are also commonly used to treat MDD, and combined with antidepressant therapy they can lead to higher improvement in depressive symptoms and better adherence to treatment (*106*). Other options include electroconvulsive therapy (ECT), transcranial magnetic stimulation, vagus nerve stimulation, and neurosurgical procedures (*107*–*109*).

1.6.3 Risk factors

A number of biological, psychological, and environmental risk factors influence the occurrence of MDD. Neuroticism, adverse childhood events, stressful life events, and low socioeconomic status all increase the risk of the disorder (*1*, *110*–*113*). Chronic medical conditions, female gender, and a family history of MDD also increase risk, with a heritability of approximately 37% (*1*, *114*). Substantial cultural differences exist in the prevalence and symptom presentation of depression (*93*).

1.6.4 Pathology

Central and peripheral abnormalities are reported in MDD and are tied to hypotheses of MDD. Additional pathological findings are described in the context of specific hypotheses of MDD following this section.

1.6.4.1 Central

Anatomical and blood flow abnormalities have been found in a number of brain regions in MDD, including the prefrontal and cingulate cortex, hippocampus, striatum, amygdala, and thalamus (*115*). Gray matter reductions have been observed in a number of brain regions (*116*). The left subgenual cingulate cortex, involved in regulation of emotional behaviour, shows volume reduction early in MDD, in young adults with high familial risk, and lesions in the area are associated with perturbed autonomic responses to social stimuli (*117*). Hippocampal atrophy occurs in recurrent MDD and is correlated with total duration of MDD (*118*), while reduced cortical thickness, neuronal size, and neuronal and glial densities have been found in areas of the prefrontal cortex (*119*). Findings from gene expression and proteomic studies of the brain in MDD include alterations in oligodendrocyte structure and function, synaptic neurotransmission, energy metabolism, and cell proliferation (*120*, *121*).

1.6.4.2 Peripheral

A number of peripheral disturbances have been found in MDD involving immune and inflammatory, oxidative stress, metabolic, neurotrophic, and endocrine factors. Elevated peripheral (blood/urinary) markers of oxidative stress have been found in depression together with perturbed antioxidant superoxide dismutase and glutathione activity, though these findings are inconsistent (*122*). Markers of inflammation, including erythrocyte sedimentation rate and blood levels of C-reactive protein (CRP), cytokines, and neopterin, are also higher in depression patients (*122–124*). Furthermore, interleukin (IL)-1β, IL-6, and CRP are reduced after antidepressant treatment (*125*, *126*). In drug naive MDD patients, higher peripheral leukocyte messenger ribonucleic acid (mRNA) levels of vascular endothelial growth factor (VEGF) have been found, and reductions were correlated with clinical improvement after antidepressant treatment (*127*). On the other hand, serum levels of brain-derived neurotrophic factor (BDNF) are reduced in depression and increase after antidepressant treatment (*128*). A serum proteomic study of first onset, antidepressant naive MDD patients found evidence for concurrent elevations in pro-inflammatory and oxidative stress response, HPA axis hyperactivity, and growth factor disruptions (*129*). A plasma biomarker panel has produced good sensitivity and specificity in separating MDD patients from controls and has implicated altered insulin, acute phase proteins, and matrix metalloproteinases in MDD (*130*). Another set of nine serum analytes involved in

inflammation, metabolism, and neurotrophic and HPA systems separated patients from controls with approximately 92% sensitivity and 81% specificity (*131*). A number of studies have reported insulin resistance in depression consistent with an increased risk for CAD, type 2 diabetes, stroke, and hypertension (*132*). Dysfunction of the HPA axis under stress been investigated through peripheral measurements and is discussed in the next section.

1.6.5 Hypotheses

Little is known about the aetiology and pathophysiology of MDD and other mental disorders. However, a number of hypotheses for depression have emerged and are briefly outlined below.

1.) Genetic and environmental factors are thought to interact in MDD. The inherited personality trait neuroticism is a considerable, but not the only, component of this **genetic vulnerability** (*133*). However, the heterogeneity of MDD and insufficient sample sizes have prevented discovery of strong, consistent genetic risk factors (*134*). Multiple genes may be involved in developing MDD, with each gene contributing a small effect (*115*). Epigenetic modifications may also alter gene expression and explain the high discordance rates between monozygotic twins, greater female prevalence, and other aetiological complexities (*135*).

2.) Little is known about the **impairment of neural circuitry** in MDD, including brain reward pathways and regulation of mood and emotion. The prefrontal cortex and hippocampus are implicated in the cognitive dysfunction of depression, the striatum and amygdala in anhedonia and anxiety, and the hypothalamus in neurovegetative symptoms of depression (*115*, *136*).

3.) The **monoamine hypothesis** proposes that reduced monoamine function in the brain leads to MDD. This hypothesis was generated by studying the mechanisms of antidepressant medications, which increase monoamine transmission (*136*). **Glutamatergic** and **GABAergic** neurotransmitter system disturbances may also be involved in MDD (*117*).

4.) Involvement of **neurotrophic factors** and neurogenesis in MDD is hypothesized from observations of decreased volumes of some brain regions in depression. BDNF is thought to be of particular importance and has been implicated in stress response and antidepressant action (*136*).

5.) **Dysregulation of the HPA axis** may result from chronic stress and be involved in the pathophysiology of MDD. Excesses in the levels of cortisol have been found in MDD patients at baseline and during recovery from stress (*136–139*). Metabolic abnormalities and hippocampal volume reductions in MDD may be partly explained by excess glucocorticoid (*140*). However, HPA axis dysfunction may depend on MDD type. Hyperactivity of the axis is associated with severe episodes, psychotic symptoms, and melancholic depression, while hypoactivity is associated with atypical depression, characterized by lethargy and hypersomnia (*141*).

6.) The **cytokine hypothesis** of depression posits that peripheral proinflammatory cytokines are responsible for the behavioural, neuroendocrine, and neurochemical changes in MDD (*142*, *143*). Cytokines activate the HPA axis and may contribute to dysregulation of the axis. Mechanisms have been identified that allow peripheral cytokines to infiltrate the brain through the blood-brain-barrier (BBB) or to indirectly signal the brain, resulting in production of cytokines by migroglia (*142*, *143*). Systemic administration of cytokines affects monoaminergic neurotransmission in animals and induces ‘sickness behaviour’ similar to depression (*136*, *142*, *143*). Alternatively, chronic psychological stress may impair glucocorticoid regulation of the immune system and result in higher cytokine levels (*144*). A recent study demonstrated that microglial function is disturbed in chronic stress-induced depression in rodents (*145*).

7.) The **vascular depression hypothesis** posits that cerebrovascular disease is linked to late-life depression (*146*). This subtype has been characterized by executive dysfunction and/or white matter hyperintensities and has a poor response to antidepressant treatment (*147*).

8.) Diurnal changes in symptoms, antidepressant effects on circadian rhythms, and shorter REM latency in MDD patients suggest that abnormalities in **circadian rhythm** may be associated with MDD. However, genetic and molecular evidence for this hypothesis is lacking (*117*).

Hypotheses of depression continue to evolve to account for and encompass all pathological findings.

1.7 Sex and gender differences in MDD

Women have a higher prevalence of MDD, a phenomenon that has been observed consistently across several countries and cultural settings from late adolescence onwards (*148*, *149*). Puberty marks the emergence of this trend, with childhood depression equally prevalent in boys and girls (*150*). In the United States, lifetime prevalence is estimated to be 17% for women and 9% for men, with similar rates observed in Europe (*151*, *152*). Suicidal behaviour shows considerable differences between men and women (*14*), with higher suicide rates for men in most countries (China being a notable exception), and trends toward further increases in this difference. On the other hand, attempted suicide and deliberate self-harm is more common in women (*14*). Women and men may also differ in their presentation, aspects of course of the disorder, and response to treatment, but there is conflicting evidence for this (*153*).

Anxiety disorders and comorbid MDD and anxiety disorder(s) (CMA) also have a higher prevalence in women (*154*). Data from the Collaborative Psychiatric Epidemiology Studies (2002-2003) (*154*) in the United States revealed anxiety disorders affected 28% of individuals, with higher female lifetime (odds ratio 1.7) and past-year prevalence rates (odds ratio 1.8) and a higher comorbidity with MDD in women with lifetime anxiety (odds ratio 1.38). This was concurrent with a greater illness burden for women with anxiety disorders. *Simonds and Whiffen (2003)* (*155*) also assessed seven studies (published 1990-1997) and concluded that women had a higher prevalence of pure anxiety or depressive and comorbid disorders. On the other hand, the prevalence of bipolar disorder has not been found to differ between men and women (*149*).

1.7.1 Hypotheses and pathology

Current explanations proposed for the marked gender difference in prevalence of MDD include artefactual and genuine biological, psychological, and social determinants (*156*, *157*). These will be discussed in more detail in this section.

1.7.1.1 Artefactual hypotheses

A greater female tendency to report more and different depressive symptoms, seek help for depression more frequently, and recall more past depressive events than men have been hypothesized as artefacts explaining the greater preponderance of women with MDD (*157*). Mixed evidence has been reported for these hypotheses. However, two large multicentre studies showed comparable gender differences in prevalence of MDD in the community and primary care setting and studies have shown no evidence for gender differences in recalling severity of depressive symptoms (*156*, *157*). Men and women may differ in types of

symptoms expressed in MDD, but this does not fully account for the difference in prevalence (*157*).

1.7.1.2 Psychological and social hypotheses

The roles of childhood abuse and adverse experiences, social roles and cultural norms, adverse life events, personality traits and coping strategies, prior anxiety, and others have been explored. **Section 1.3** outlines gender differences in some of these factors. Mixed evidence has been presented for social and other psychological factors causing more MDD in women, including chronic strain resulting from female social status, role limitation, use of rumination as a coping strategy, and greater vulnerability and/or sensitivity to different stressors, life events, and major traumas, in particular sexual abuse (*156–158*).

1.7.1.3 Biological hypotheses

1.7.1.3.1 Sex hormones

Sex hormones have been implicated in the emergence of higher rates of female depression during puberty and rising hormonal levels during this time have been linked to affective disturbances in girls (*156*). Hormones may influence gene expression and interact with negative life events to produce the gender difference in MDD prevalence at puberty (*159*). Fluctuating hormone levels associated with the female cycle and other reproductive events, such as pregnancy and menopause, cause disturbances in neurochemical pathways (*160*). A cycle-dependent vulnerability for affective symptoms in the premenstrual phase has been found in approximately 1% of women (*156*). The postpartum period is a period of increased risk for depression and may be associated with the abrupt change in hormone levels at this time, though rates vary considerably across cultures (*160*). Similarly, the incidence of depression in women is higher during perimenopause (*160*).

1.7.1.3.2 HPA axis dysfunction

Animal studies of HPA axis activity have found lower basal and stress-induced glucocorticoid and adrenocorticotropic hormone (ACTH) levels in males, but these findings have not been upheld in human studies (*29*, *161*). Most acute real-life and laboratory studies of HPA axis response to psychological stress have found no sex differences or a higher cortisol response in young men, but results vary with stressor protocol, population studied, female hormonal status, and other factors (*161*). Females have been reported to have greater cortisol responses after social rejection, while males were found to have greater cortisol responses to achievement challenges (*162*). Cortisol response to stress is affected by female hormonal status, with smaller responses to the Trier Social Stress Test in the follicular phase compared to the luteal phase and severely blunted responses in OC users (*158*).

Menopause is associated with increased dexamethasone resistance compared to premenopausal women (*158*). Some evidence suggests that depressed females may show greater HPA axis dysregulation (*161*, *163*, *164*), while others have found the reverse, with higher baseline levels of salivary cortisol only in medication-free depressed males (*165*) and in adolescent males who later developed clinical depression (*166*). *Goldstein et al (2014)* (*164*) suggested study inconsistencies may be due to the timing of cortisol measurement, patient age and hormonal status, genetic factors, chronicity of MDD, or low statistical power.

1.7.1.3.3 Inflammation and immune function

Recent studies of depression have found sex-dependent differences in inflammation. Several large studies have found an association between elevated levels of CRP and depression in males only, though this finding is not entirely consistent (*124*, *167–172*). Furthermore, a 2009 meta-analysis of inflammatory markers in depression showed that samples consisting only of men showed a significant increase in CRP while samples of only women did not (*124*). However, IL-1 and -6 were also increased in depression with no evidence of sex differences, while another large meta-analysis of cytokine levels in depression did not investigate sex differences (*123*, *124*). A study by *Toker et al (2005)* (*169*) also found that the association between burnout, depression, and anxiety and levels of CRP and fibrinogen varied in a sex-specific manner across these states and speculated that stress pathways may influence inflammatory processes differently in males and females.

1.7.1.3.4 Other hypotheses

Other hypotheses implicate genetic factors, the hypothalamic-pituitary-thyroid (HPT) axis, and aspects of brain function (*156*, *157*). A genotype-stress interaction involving 5-HTTLPR in MDD has been demonstrated for female adolescents only and there is evidence that MAOA polymorphisms on the X-chromosome produce more vulnerability to depression in females (*159*). Skewed (non-random) X-chromosome inactivation has also been proposed as an epigenetic process that could contribute to higher rates of MDD in females (*135*). Alternatively, affective disorders in subgroups of females may be caused by subclinical hypothyroidism, which is associated with reduced central serotonergic activity (*156*). However, subsequent studies have failed to confirm this. Sex differences in brain development and function, discussed in **Section 1.2.1,** are also hypothesized to play a role (*159*).

1.7.1.4 Integrated models

No single explanation for the preponderance of depression in females has been found. Affective and cognitive factors are thought to interact with biological factors to contribute to

the greater female prevalence of MDD (*157*, *159*). Models integrating these factors to influence the female preponderance in MDD have been proposed, as seen in **Figure 1.7**.

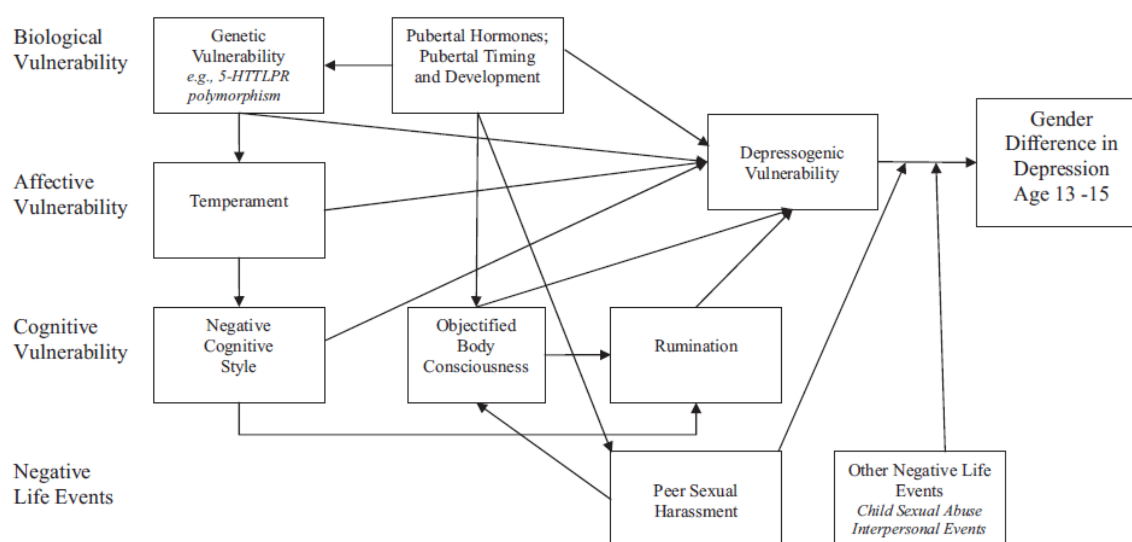


Figure 1.7. An affective, biological, and cognitive model explaining gender differences in prevalence for MDD from *Hyde et al (2008)* (*159*). *Reprinted with permission.*

1.8 Schizophrenia

Schizophrenia is a chronic, severe psychotic disorder with heterogeneous symptoms impairing cognition, behaviour, emotions, and social and occupational functioning. Symptoms are positive (hallucinations, delusions, thought and movement disorders) and negative (flat affect, poverty of speech, avolition, anhedonia, asociality). Cognitive deficits (impaired executive functioning, working memory, attention) are also common. Schizophrenia has a lifetime prevalence of approximately 0.7% (*173*). Though rare, it is a severely debilitating and costly disorder with poor outcome. Sufferers are often incarcerated, homeless, or unemployed (*92*) and their lifespan is shortened by 13-30 years (*174*). Schizophrenia is the 5th and 6th leading global cause of years lost to disability for men and women, respectively, and 8th leading cause of DALYs in the 15-44 years age group (*175*, *176*). Psychotic disorders account for annual estimated direct and indirect costs of €94 billion in Europe and are considered the most expensive mental illness per patient treated (*96*, *177*).

Other mental disorders may be associated with psychosis or share similar symptoms with schizophrenia. These include schizoaffective, schizophreniform, brief psychotic, schizotypal personality, delusional, obsessive-compulsive, body dysmorphic, posttraumatic stress, and

autism spectrum disorders. MDD, bipolar disorder, and other mental disorder may also have psychotic or catatonic features. Substance abuse is prevalent in schizophrenia, both before and after diagnosis (*178*, *179*), and this is associated with young age, earlier onset, violent crime, less education, male gender, and a worse outcome in patients (*178–180*). Tobacco smoking is also more common in schizophrenia patients, particularly in males (*181*). Physical illness is more common in schizophrenia, due in part to lifestyle and antipsychotic side effects. Schizophrenia carries an increased risk for the metabolic syndrome together with cardiovascular disease and type 2 diabetes (*182*).

1.8.1 Criteria for diagnosis

The DSM-IV criteria for diagnosis of schizophrenia (*22*) are again presented in **Table 1.3**, as this work and most other published research has not been carried out with the most recent 5th edition.

Table 1.3. DSM-IV criteria for diagnosis of schizophrenia.

A. Two (or more) of the following, each present for a significant portion of time during a 1-month period: a. Delusions b. Hallucinations c. Disorganized speech d. Grossly disorganized or catatonic behaviour e. Negative symptoms **Note:** Only one Criterion A symptom is required if delusions are bizarre or hallucinations consist of a voice keeping up a running commentary on the person's behaviour or thoughts, or two or more voices conversing with each other.
B. For a significant portion of the time since the onset of the disturbance, one or more major areas of functioning such as work, interpersonal relations, or self-care are markedly below the level achieved prior to the onset.
C. Continuous signs of the disturbance persist for at least 6 months. This 6-month period must include at least 1 month of symptoms that meet Criterion A and may include periods of prodromal or residual symptoms.
D. Schizoaffective Disorder and Mood Disorder With Psychotic Features have been ruled out because either (1) no Major Depressive Episode, Manic Episode, or Mixed Episode have occurred concurrently with the active-phase symptoms; or (2) if mood episodes have occurred during active-phase symptoms, their total duration has been brief relative to the duration of the active and residual periods.
E. The disturbance is not due to the direct physiological effects of a substance or a general medical condition.
F. If there is a history of Autistic Disorder or another Pervasive Developmental Disorder, the additional diagnosis of Schizophrenia is made only if prominent delusions or hallucinations are also present for at least a month.

Five subtypes of schizophrenia are defined in the DSM-IV based on the predominant symptomatology at clinical presentation. These include paranoid, disorganized, catatonic, undifferentiated, and residual types.

Changes in the DSM-5 include removal of the exception to Criterion A requiring only one symptom for bizarre delusions or hallucinations of voices (*1*). However, one of the core positive symptoms (a, b, or c in Criterion A) is still required for a diagnosis of schizophrenia. Subtypes of schizophrenia were also removed because they were found to inadequately represent the heterogeneity of the illness, showing instability over illness course, lack of clinical value, and poor reliability and validity (*183*). These have been replaced with a rating scale for the severity of the core symptoms (*1*).

1.8.2 Course and treatment

Psychotic symptoms of schizophrenia usually manifest between late adolescence and the mid-30s, while childhood and late-onset (after 40 years) are rarer. Onset of psychosis is most often preceded by signs and symptoms in the prodromal stage, including poor functioning, low-grade psychotic symptoms, depression and anxiety, and disorganization (*184*). A premorbid phase may occur prior to this prodromal phase in which early-stage dysfunction manifests. The course of schizophrenia is highly variable, favourable in approximately 20% of patients, with the majority requiring substantial daily support, experiencing relapses and remission of psychotic symptoms, and some progressively deteriorating (*1*). About 5% of schizophrenia patients commit suicide and more will attempt suicide or have suicidal ideation (*185*). While the psychotic symptoms of schizophrenia are generally reduced over the course of illness, cognitive and negative symptoms persist. Late-onset schizophrenia has more positive symptoms, less family history, and a higher female incidence (*186*).

From the mid-1950s, antipsychotic medication has been used for treatment and maintenance of positive symptoms in schizophrenia. First-generation antipsychotics (FGAs) include chlorpromazine, haloperidol, and perphenazine and act by blocking dopamine D_2 receptors (*187*). However, FGAs cause extrapyramidal symptoms and second-generation antipsychotics (SGAs) were introduced in the 1980s. SGAs include clozapine, olanzapine, quetiapine, and risperidone, and block both dopamine D_2 and serotonin 2A (5-HT_{2A}) receptors (*187*). These SGAs increase the risk of metabolic side effects, including weight gain, hyperglycemia, dyslipidemia, and type 2 diabetes. FGAs and SGAs have not been found to differ in their efficacy or tolerability (*188*). However, lack of response to antipsychotics, poor adherence, and lack of effective treatment for negative symptoms and

cognitive deficits continue to cause severe impairment in schizophrenia. Psychological and psychosocial interventions are also used to improve outcome in schizophrenia (*189*).

1.8.3 Risk factors

A number of biological, psychological, and environmental risk factors influence the occurrence of schizophrenia. Babies born in late winter and early spring in the Northern Hemisphere, to older fathers, and who are exposed to hypoxia and maternal infection, stress, and malnutrition are more prone to later develop schizophrenia (*190*, *191*). Children diagnosed with schizophrenia in adulthood are also more likely to have poor social functioning, emotional and behavioural disturbances, intellectual and language impairments, and motor delays (*191*). Growing up in an urban environment, immigrant status in certain ethnic groups, male gender, cannabis use, and social disadvantage are also associated with a higher risk for schizophrenia (*173*, *191*). Family history increases the risk of schizophrenia and it has a high heritability of approximately 80% (*191*). However, prevalence varies considerably between sites and diagnostic criteria affects incidence estimates differently in men and women (*173*, *191*).

1.8.4 Pathology

Both central and peripheral abnormalities have been found in schizophrenia and are discussed below. Studies of the pathology of schizophrenia and hypotheses for its aetiology and pathophysiology are intimately linked. A number of additional pathological findings are also discussed in **Section 1.8.5** in the context of specific hypotheses of schizophrenia.

1.8.4.1 Central

Studies have found structural and functional brain abnormalities in schizophrenia together with altered gene and protein expression. A large meta-analysis of structural magnetic resonance imaging (MRI) studies found that the mean volume of the cerebrum was decreased (98% of control volume) and ventricular volume was increased (126%) in schizophrenia (*192*). Decreased amygdala, hippocampus, and parahippocampus volumes were found in excess of the global difference. This may be due to decreased size of neuropils and neurons (*193*). Reduced gray and white matter volume and increased sulcal and ventricular cerebrospinal fluid (CSF) have also been reported in schizophrenia (*194*). A number of these abnormalities have been observed in first onset, unmedicated patients, though they may progress over time (*195*) and be influenced by long-term antipsychotic use (*196*). Functional imaging data points to dysfunction of the activity and integration of circuits in the prefrontal cortex, hippocampus, and subcortical structures (*193*). The dorsolateral prefrontal cortex (DLPFC) is thought to be important to the cognitive deficits in schizophrenia and a number of abnormalities have been found in this region, including reduced metabolism

and blood flow and other disruptions in neurotransmission and cortical connections (*197*). Studies of gene expression and proteomic investigations have found a number of biological processes altered in schizophrenia post-mortem brains, including oligodendrocyte function, energy metabolism, oxidative stress, myelination, inflammation, synaptic transmission, cell communication, and signal transduction (*51*, *198–204*).

1.8.4.2 Peripheral

Peripheral metabolic, neurotrophic, endocrine, and immune abnormalities have been found in schizophrenia and many of these may reflect central disturbances. Altered lipid metabolism has been found, with evidence for an association between peripheral membrane abnormalities and perturbed brain phospholipid metabolism in schizophrenia (*89*, *205*). Additionally, eicosapentaenoic acid may be beneficial in schizophrenia and omega-3 fatty acids are reported to reduce conversion to psychosis in high-risk individuals (*89*, *206*). Impaired fasting glucose tolerance and more insulin resistance has also been reported in first-episode, drug naive schizophrenia patients (*207*). Cerebral glucose metabolism abnormalities have been associated with peripheral platelet mitochondrial complex I activity (*208*). Furthermore, increased serum insulin, chromogranin A, pancreatic polypeptide, prolactin, progesterone, and cortisol levels, and decreased growth hormone levels found concurrently in first and recent onset patients have implications for dysfunction of neuroendocrine systems (*209*). Reduced blood concentration of BDNF (*210*) has been reported in schizophrenia and elevated levels of inflammatory markers in blood are a robust finding (*211*) in the disorder. A biological signature for schizophrenia comprised of ten plasma analytes has distinguished schizophrenia patients from controls with a sensitivity and specificity of over 90% and 95%, respectively (*130*), while another panel of 51 serum inflammatory and metabolic molecules, hormones, and others has yielded 83% sensitivity and 83% specificity in separating patients from controls in an independent cohort (*212*). This multiplex immunoassay panel has also been used to identify 21 analytes differentially expressed in schizophrenia in post-mortem brain tissue, nine of which were altered at the transcriptional level in the same patients, and another nine of which were reproducibly altered in serum of first onset patients (*213*). Another approach has uncovered six genes differentially expressed in both blood and brain (*214*).

1.8.5 Hypotheses

Remarkably little is known about the aetiology and pathophysiology of schizophrenia. Nevertheless, current hypotheses of schizophrenia are briefly outlined below. Many of these hypotheses are not necessarily mutually exclusive and may work at different levels.

1.) Schizophrenia is highly heritable and it is hypothesized that **genetic vulnerability** interacts with environmental factors to influence its development. Multiple genes are thought to be involved, each conferring a small risk (*215*). A 2014 meta-analysis of genome-wide association studies (GWAS) identified 128 associations in schizophrenia for 108 genetic loci, including the dopamine D_2 receptor and genes involved in glutamatergic neurotransmission and synaptic plasticity (*216*). Findings were enriched for genes expressed in brain and tissues important for immunity. In particular, strong enrichment was found for the mid-frontal lobe, angular gyrus, and in B-lymphocyte lineages (CD19 and CD20 lines). Altered gene expression in schizophrenia is implicated from these GWAS results, rather than altered protein structure. Copy number variants may account for 2% of cases (*217*) and epigenetic factors may play a role in schizophrenia (*218*).

2.) **Neurotransmitter dysfunction** has long been implicated in schizophrenia.

- **a.** The **dopamine hypothesis** proposed that schizophrenia was due to an excess of dopamine in the brain after it was found that FGAs blocked dopamine D_2 receptors. Subsequent findings lead to a revision of the hypothesis and proposed that prefrontal hypodopaminergia results in subcortical hyperdopaminergia to cause schizophrenia. The current hypothesis posits that presynaptic dopamine dysregulation is the final common pathway to psychosis in schizophrenia (*219*).
- **b.** Evidence of **NMDA receptor (NMDAR) hypofunction** in schizophrenia has come from animal and clinical studies (*220*). Chronic use of phencyclidine (PCP), a noncompetitve NMDAR antagonist, induces schizophrenia-like positive, negative, and cognitive symptoms. Trials with glycine modulatory site agonists on the NMDAR have yielded reduced negative symptoms of schizophrenia (*220*).
- **c.** The **GABAergic origin hypothesis** of schizophrenia pathophysiology suggests that GABAergic interneurons mediate symptoms caused by NMDAR hypofunction (*221*). Schizophrenia-like phenotypes can be generated using transgenic mice with selectively removed NMDARs from GABAergic interneurons in the cortex and hippocampus (*221*).
- **d.** The neurotransmitter **serotonin** is also implicated in schizophrenia. Chronic serotonergic overdrive in the cerebral cortex caused by stress may cause disrupted glutamatergic signalling, cortical atrophy, and peripheral phospholipid abnormalities in schizophrenia (*222*).

e. The **cholinergic system** may play a role in schizophrenia, suggested by the high prevalence of smoking in schizophrenia patients (*223*).

A **circuit-based framework** has been developed by *Lisman et al (2008)* (*223*) to model how risk genes and neurotransmitters in schizophrenia interact, and our understanding of these systems will no doubt continue to evolve.

3.) A number of observations, including the increased risk for schizophrenia after maternal infection and elevated peripheral and central markers of inflammation, have lead to **immune hypotheses** of schizophrenia, including the macrophage-T-lymphocyte (*224*, *225*), cytokine (*226*), microglia (*227*), Th2 (*228*), fetal brain cytokine imbalance (*229*), genetic-inflammatory-vascular (*230*), and mild encephalitis (*231*) hypotheses of schizophrenia. A 2011 meta-analysis (*211*) of blood levels of cytokines found support for the macrophage-T-lymphocyte and microglia hypotheses of schizophrenia. The macrophage-T-lymphocyte theory of schizophrenia (*224*, *225*) proposes that chronically activated macrophages and T-lymphocytes are essential mediators of schizophrenia and suggests the gastrointestinal tract as the potential cause of immune activation. Release of cytokines and free radicals by activated microglia are thought to be important in the microglia hypothesis of schizophrenia (*227*). Findings from the meta-analysis were less consistent with the Th2-hypothesis that posits a shift to Th2-like immune reactivity in schizophrenia. Low-level neuroinflammation is an important part of the pathophysiology of schizophrenia and affective disorders according to the mild encephalitis hypothesis, caused by autoimmunity, infection, trauma, or toxicity (*231*). In this context, a higher prevalence of NMDA receptor antibodies have been found in serum of patients with an initial diagnosis of schizophrenia (*232*). Alternatively, cytokine signalling from peripheral to immature brain tissue during development may result in abnormal brain development in schizophrenia (cytokine hypothesis) (*226*) or elevated maternal cytokines may influence neurodevelopment through an imbalance in pro- and anti-inflammatory cytokine signalling in the fetal brain (fetal brain cytokine imbalance hypothesis) (*229*). A genetic-inflammatory-vascular theory of schizophrenia further proposes that a genetic predisposition to exaggerated inflammatory response damages the brain's microvascular system, causing metabolic disturbances (*230*). Recent GWAS results discussed above were enriched for genes expressed in tissues important for immunity, providing further evidence for involvement of the immune system in schizophrenia.

4.) The **diathesis-stress model** proposes that schizophrenia is caused by a biological vulnerability together with exposure to stress. Stressful life events can increase the severity of schizophrenia symptoms (*233*) and higher post-dexamethasone suppression test cortisol levels are associated with poor prognosis, greater premorbid impairment, and ventricular enlargement in schizophrenia (*233*).

5.) The **neurodevelopmental hypothesis** of schizophrenia posits that abnormalities in brain development cause schizophrenia in adulthood and is illustrated in **Figure 1.8**. The **two-hit model** further proposes that these abnormalities occur during critical developmental time points early in life and during adolescence. This hypothesis is supported by genetic and early risk factors that affect development, congenital anomalies, brain pathology (*234*), and disturbed **neurotrophic factors** (*235*) in schizophrenia.

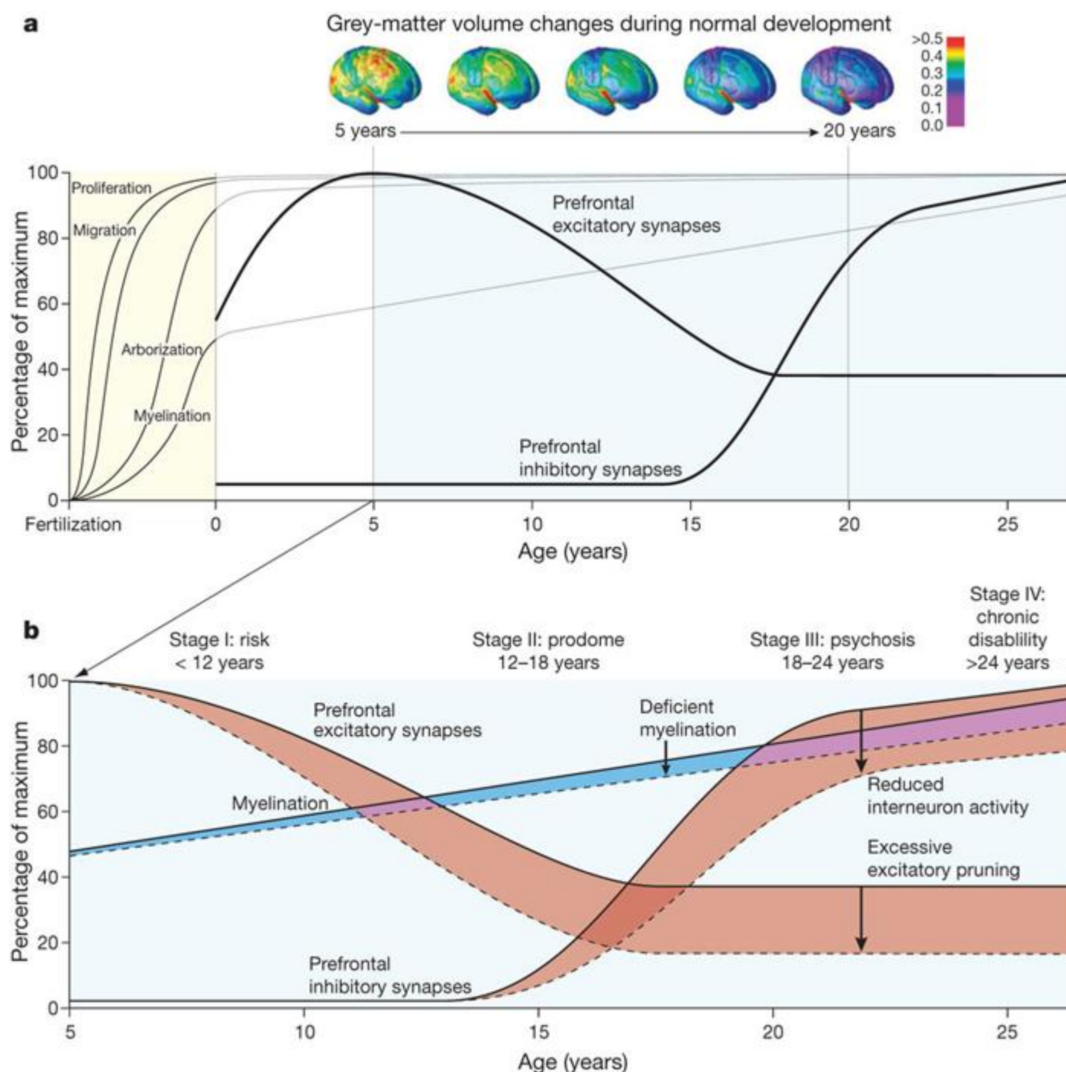


Figure 1.8. Neurodevelopmental model of schizophrenia; from *Insel (2010)* (*92*)**. (A)** illustrates normal cortical developmental processes, while **(B)** illustrates neurodevelopmental processes that may occur in schizophrenia. *Reprinted with permission.*

As with MDD, hypotheses of schizophrenia are constantly developing in order to account for all pathological findings.

1.9 Sex and gender differences in schizophrenia

A higher male incidence of schizophrenia has been found in a number of recent studies, despite a similar prevalence in males and females (*12*, *236*). A risk ratio of 1.4:1 for male compared to female incidence was reported in a 2003 meta-analysis, with more restrictive definitions of schizophrenia yielding higher ratios (*237*). Several studies have shown that males have more negative symptoms while females have more affective symptoms (*238*), although some have not found sex differences in symptomatology (*236*). Females have also been found to suffer from more atypical and cyclical psychoses, which together with more affective symptoms reduce diagnostic concordance for women under different systems (*238*).

Studies have also consistently shown that females typically have an older age of schizophrenia onset (*12*, *236*). A mean age of onset of 30.1 years in females and 26.7 years in men was found in the WHO Determinants of Outcome Study (*239*), which began in the 1980s and examined patients aged 15-54 from centres in 10 countries. Whereas the male age-at-onset distribution showed a single peak in the early twenties across countries, the female distribution showed a broad peak in the twenties with a second peak after age 40, illustrated in **Figure 1.9** (*239*). Similar results were found in the ABC Schizophrenia Study (1987-1989) in Germany (*239*). Evidence for an additional female peak after age 60 was also found in a study by *Castle, Sham, and Murray (1998)* (*240*) that included older subjects. Sex differences in age of onset are consistent across cultures, definitions of onset, and diagnostic criteria of illness, and cannot be accounted for by gender differences in patient or family help-seeking behaviour (*238*). Studies have shown that the older female age of onset may occur only in the paranoid subtype of schizophrenia (*241*, *242*) and that it dissipates in patients with a family history of schizophrenia (*238*). Together with better premorbid functioning, the older female age of onset is thought to account for the better social course and short- to medium-term prognosis of illness in women compared to men (*12*, *236*, *243*, *244*). Greater rates of substance abuse are present in men with schizophrenia, which are associated with poor adherence to treatment, more severe psychotic symptoms, violence, and homelessness (*179*, *236*). Premenopausal women have also been reported to respond better to treatment with antipsychotic medication and require lower doses compared to their male counterparts (*243*, *245*, *246*). Postmenopausal women require higher antipsychotic doses and may develop more severe symptoms (*12*, *243*).

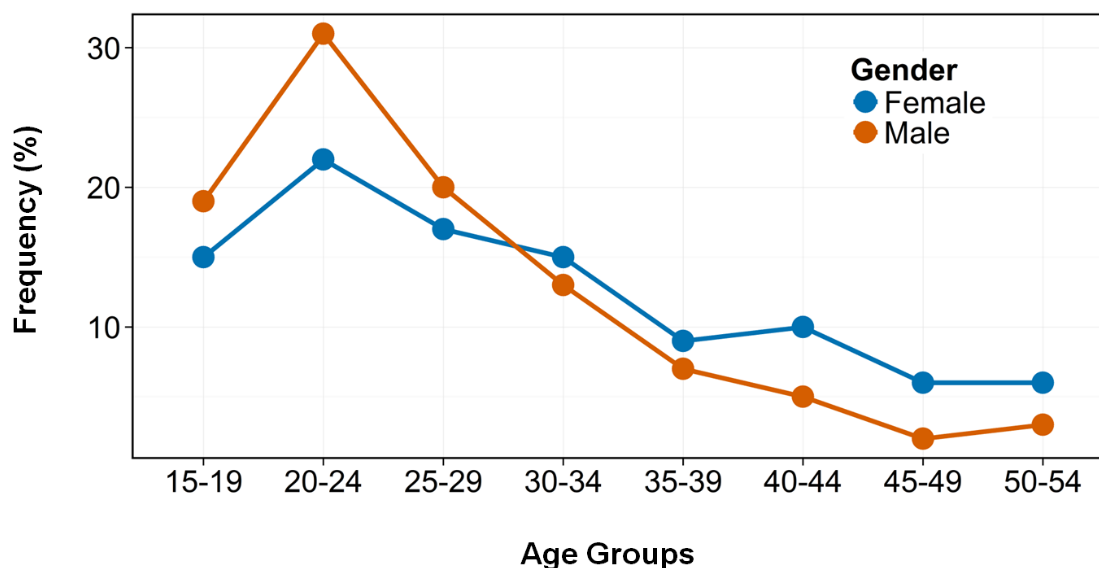


Figure 1.9. Male and female age-at-onset distributions for schizophrenia in the WHO Determinants of Outcome 10-country Study (*239*)**.**

1.9.1 Pathology

Despite the evidence for sex-based heterogeneity in schizophrenia, many studies do not report whether the results of analyses differed for males and females. Nevertheless, sex differences in brain abnormalities, HPG axis involvement, and genetic risks in schizophrenia provide evidence for differences in aetiology and/or pathophysiology between males and females. In general, a greater number of structural brain abnormalities have been found in male compared to female schizophrenia patients, including enlarged ventricles and smaller temporal and frontal lobes (*247*), though a meta-analysis in 2000 of regional brain volumes did not find evidence for this (*248*). A MRI study conducted by *Goldstein et al (2002)* (*249*) found disruptions in normal sexual dimorphisms in cortical brain volumes in schizophrenia, including sex-specific abnormalities in frontomedial cortex, basal forebrain, and planum temporale volumes and planym temporale asymmetry. Similarly, the orbitofrontal cortex to amygdala ratio was reported to be feminised in males with schizophrenia and masculinised in females (*250*). Other functional MRI (fMRI) studies have found reversals of normal sexual dimorphisms in cerebral activation during processing of aversive stimuli (*251*) and during mental rotation (*252*). A more prominent increase in hypothalamic volume has been found in female schizophrenia patients and non-psychotic first-degree relatives compared to controls (*253*). The magnitude of overall and region-specific reduced gray matter volume and their correlation with clinical variables in schizophrenia may be mediated by sex (*254–257*). Sex-specific differences in cerebral asymmetry and craniofacial shape asymmetries may also occur in schizophrenia (*258*, *259*). However, some of these findings are inconsistent and many require replication (*238*).

A number of studies have investigated the role of estrogen in schizophrenia. Psychosis has been associated with reduced estrogen levels after menopause and postpartum (*260*), while pregnant females experience fewer relapses and symptoms of schizophrenia are improved during the higher estrogen phase of the menstrual cycle (*260–263*). Hyperprolactinemia induced by typical and some atypical antipsychotics (*264*) is thought to contribute to menstrual disturbances, sexual dysfunction, and reduced fertility in female patients (*265*). However, studies dating from before the introduction of antipsychotic medication have found hypoestrogenism in female schizophrenia patients (*266*) and more recent studies have confirmed these findings in all phases of the menstrual cycle in patients who were using both typical and prolactin-sparing atypical antipsychotic medication (*263*, *267*). A 2012 meta-analysis concluded that estrogen treatment improved total, positive, and negative symptom severity in females with schizophrenia (*9*). Evidence of further gonadal dysfunction in schizophrenia has been found, though findings vary with chronicity of illness and treatment with antipsychotic medication. In treated patients, *Bergemann et al (2005)* (*267*) found low serum LH and progesterone and high testosterone levels. On the other hand, studies of male and female recent and first onset, antipsychotic naive schizophrenia patients have found elevated serum progesterone, prolactin, FSH, and LH (*209*, *268*). Studies of testosterone and other androgens in men with schizophrenia have produced conflicting results, but some have found an inverse correlation between testosterone levels and negative symptom severity (*269*, *270*).

Sex-specific and sex-dependent genetic risks may contribute to the heterogeneity of schizophrenia in males and females, but these require confirmation (*271*). These include genes with roles in dopaminergic metabolism catechol-O-methyltransferase (COMT) and monoamine oxidase (MAO) (*272*). Furthermore, three single-nucleotide polymorphisms (SNPs) on the X-chromosome were associated with schizophrenia in the most recent GWAS meta-analysis (*216*).

1.9.2 Hypotheses

A number of hypotheses have been proposed to explain the sex-based heterogeneity in schizophrenia, and these are described below.

1.) Differences between male and female schizophrenia patients have lead to speculation that the **sexes are differentially predisposed to subtypes of the disorder** (*273–276*). Males may be more prone to schizophrenia caused by early environmental insults with a neurodevelopmental basis of illness. The late-onset form of schizophrenia more common in females may be aetiologically more similar to affective psychosis.

2.) Several **neurodevelopmental theories** of sex differences have been proposed.

   a. An **early neurodevelopmental model** of sex differences proposes that sex steroid hormones in males and females during pre- and neonatal development mediate susceptibility to brain lesions that later cause schizophrenia (*238*). Greater male brain lateralization may also reduce their ability to compensate for left brain injury.

   b. A **late neurodevelopmental model** posits that sex hormones influence excessive synaptic pruning in the prefrontal cortex that may cause schizophrenia (*238*).

   c. Neuroimaging and clinical data have lead to the hypothesis that schizophrenia is characterized by a reversed or **disturbed cerebral sexual dimorphism** (*277*). Sex hormone imbalances present during organization and activation may cause female masculinisation/un-feminisation and male feminisation/un-masculinisation of the brain in schizophrenia, leading to onset.

3.) The **estrogen hypothesis** of schizophrenia proposes that estrogen is responsible for the reduced incidence, delay in female onset, more favourable course, and better response to treatment with antipsychotic medication in premenopausal women (*260*). Clinical, epidemiological, and animal studies have provided evidence for this hypothesis. In addition to its role in the brain, summarized in part in **Section 1.2.1**, estrogen also plays an important role in immune and inflammatory, cardiovascular, and metabolic processes implicated in schizophrenia.

4.) Sex differences in genetic risk, or **gene-sex interactions**, may account for sex differences in schizophrenia (*272*). However, exploration of this topic is relatively new.

1.10 Study aims and outline

Despite much research, the aetiologies and pathophysiologies of schizophrenia and MDD have not been elucidated. New diagnostic and therapeutic technologies are needed to facilitate care of patients with these disabling conditions. Sex- and gender-based differences in incidence, prevalence, age of onset, symptoms, course of disorder, and response to antipsychotic medication have been found, the causes and mechanisms of which have also

remained elusive. The main aim of this thesis is to explore molecular sex differences and sex-dependent molecular differences in schizophrenia and MDD and their potential implications for pathophysiological mechanisms and biomarker studies and technologies. This will be accomplished by molecular profiling of serum and brain tissue using multiplex immunoassay and microarray technologies, respectively.

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

- **Chapter 2** explores the experimental and statistical methods used to obtain the findings presented in subsequent chapters. This includes an outline of the multiplex immunoassay and microarray technologies used for molecular profiling of serum and brain tissue. Data pre-processing and analysis methods for the data are also described.

- **Chapter 3** assesses variation in serum concentrations of proteins and small molecules in males and females in the follicular and luteal phases of the menstrual cycle, using OCs, and after menopause not affected by MDD or other mental disorders. It discusses the potential implications of these findings for schizophrenia and MDD and assesses the importance of recording sex and hormonal status for future biomarker studies.

- **Chapter 4** investigates sex differences in serum biomarkers of MDD and their potential relevance to diagnostic technologies. It tests whether these sex-dependent markers are present in CMA and persist in remitted MDD. Finally, it explores HPG axis function in females with current and remitted MDD and CMA compared to controls by assessing variations in the serum concentrations of reproductive hormones across follicular and luteal phases of the menstrual cycle, OC use, and after menopause.

- **Chapter 5** explores differences in the levels of serum proteins and small molecules between first onset, antipsychotic naive schizophrenia patients and controls and their modification by sex. Sex differences in the associations of these serum molecules with positive, negative, and general symptoms of schizophrenia are further investigated together with changes in their concentrations after six weeks of treatment with antipsychotic medication.

- Variations in gene expression in post-mortem prefrontal cortex tissue with schizophrenia diagnosis, sex, and sex-diagnosis interactions are explored in **Chapter 6.** These are compared to serum findings from **Chapters 3** and **5**. Potential confounding with suicide status and lifetime antipsychotic, alcohol, and drug use are also evaluated.

- Finally, **Chapter 7** provides an integrated summary of the findings from **Chapters 3-6** and their implications, discusses limitations of the work in this thesis, and provides an outlook on future studies needed to improve and build upon this research.

Chapter 2 Methods

2.1. Human DiscoveryMAP® multiplex immunoassay

The Human DiscoveryMAP® is a multiplex immunoassay from Myriad-RBM (Austin, TX, USA) that measures high and low abundance proteins and small molecules. It uses Luminex xMAP technology, which combines multiplexed microsphere-format enzyme-linked immunosorbent assays (ELISAs) with flow cytometry. Polystyrene microspheres measuring 5.6 µm are coated in a primary antibody to capture a specific protein of interest. Different types of antibodies are used to coat the microspheres, each of which can be identified by an internal dye composed of a unique ratio of red to infrared fluorophores. A mixture of antibody-coated microspheres is combined with serum in a 96-well plate and incubated for one hour, during which time target proteins/molecules bind to the antibodies coating the microsphere. Detection antibodies, which are labelled with a fluorescent dye reporter signal, are then added and incubated for another hour. They bind to a specific part of the protein of interest and a flow cytometer then funnels the microspheres single-file into a detection chamber, where they are identified and quantified. A red laser is used to identify the microsphere by exciting the internal dye, while a green laser excites the fluorescent dye on the detection antibodies to quantify the amount bound. Multiplexing capabilities come from the use of flow cytometry and the unique internal dyes identifying each type of antibody-coated microsphere. An overview of this process is shown in **Figure 2.1.**

In this work, the protocol for the study participants, collection and storage of clinical samples, and test methods were carried out in compliance with the Standards for Reporting of Diagnostic Accuracy (STARD) initiative. Blood samples were collected following strict standard operating procedures in the morning, with no systematic differences between patients and controls. Venous puncture into S-Monovette 7.5 mL serum tubes (Sarstedt; Numbrecht, Germany) was used, and serum was prepared using standard protocols, leaving samples at room temperature for two hours to allow blood to coagulate. They were then centrifuged for five minutes at 4,000×g to remove particulate material and supernatants and stored at –80 ^{o}C in LoBind Eppendorf tubes (Hamburg, Germany). Samples (250 µL) were shipped on dry ice and the assay procedure was carried out in a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory at Myriad-RBM (Austin, TX, USA). Samples were blinded to analysts until all assays are completed. Assays were calibrated using duplicate 8-point standard curves and raw intensity measurements were converted to absolute protein concentrations using proprietary software. Performance of the instrument was verified using low, medium, and high concentration quality control samples for each

analyte. Concentrations were reported as ng/mL, pg/mL or mIU/mL, as appropriate. Samples with insufficient volumes and analytes above or below the concentrations on the calibration curves were labelled accordingly.

The panel of proteins and molecules was selected based on prior knowledge of peripheral biomarkers of disease and is not specific to mental disorders, constituting a semi-hypothesis driven approach. The panel focuses largely on metabolic and immune response functions and includes cytokines, chemokines, hormones, growth factors, transport proteins, acute phase response proteins, and other analytes important in mental disorders and other diseases. The specific molecules assayed vary between versions of the DiscoveryMAP®. The 95/171 analytes from the panels analysed in this work are listed in the **Appendix** (**Table A.1**). Measured proteins and steroid hormones important in reproductive function include testosterone, prolactin, FSH, LH, sex hormone-binding globulin (SHBG) and progesterone. The DiscoveryMAP® has been used to explore molecular changes in mental disorders, cancer and autoimmune, cardiovascular, and various other diseases (*10*, *278–284*).

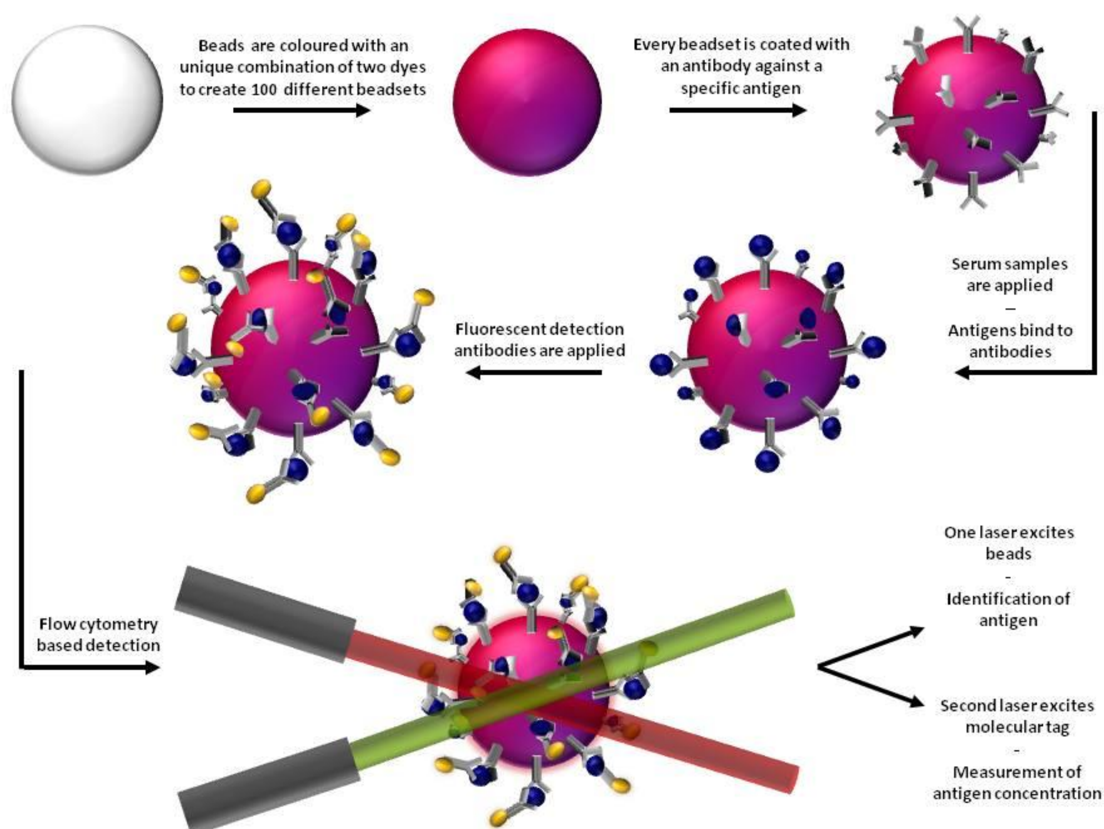


Figure 2.1. **Overview of Luminex xMAP technology used to quantify serum concentration of proteins and small molecules.** *Image courtesy of Viktoria Stelzhammer; reprinted with permission.*

2.1.1 Advantages

Proteomic methods provide an advantage over gene expression by directly measuring the functional components of cells. This results in a more straightforward determination of dysfunction in cellular mechanisms that underlie mental disorders. Compared to other technologies, the high dynamic range and throughput are key advantages of the multiplex immunoassay platform. These advantages reduce measurement variability and deterioration caused by lengthy individual experiments and allow simultaneous measurement of high and low abundance proteins. Furthermore, the required sample volume is relatively low, enabling valuable samples to be conserved.

2.1.2 Disadvantages

Incorporating new assays into the platform is expensive and costly. Therefore, relatively few proteins and molecules are measured and these are biased toward particular biological pathways and functions. Optimising signal intensity in the multiplex immunoassay requires compromise to minimise cross reactivity of detection antibodies with non-specific antigens and balance dilution ranges. Additionally, assays do not distinguish between post-translational modifications, alternative splicing, and cleaved proteins. Plate and batch effects can also cause large-scale variation in concentration measurements. These factors need to be taken into consideration when analysing and interpreting data from this multiplex immunoassay platform.

2.1.3 Clinical samples

Psychiatric diagnoses were made as described in individual chapters. Recruitment, inclusion and exclusion criteria, and demographic information are also listed in separate chapters. The ethical committees of all participating universities approved the protocols of the studies and informed written consent was obtained. Where it was required, local ethical approval for use of human serum samples was granted by the Cambridgeshire Local Research Ethics Committee. Further details about the clinical samples can be found in individual **Chapters 3-6**. Differences in demographic and other variables between groups to be compared in each chapter were assessed using Fisher's exact test for categorical variables and a *t*-test or analysis of variance (ANOVA) for continuous variables using the statistical programming software R (v3.1.2) (*285*) and $p<0.05$ to indicate significance.

2.1.4 Multiplex immunoassay data

Data pre-processing for all samples was first carried out to remove assays with more than 30% missing values using R. Lists of analytes present after this step, their abbreviations and UniProtKB accession or PubChem compound identifiers, and the percentage of missing

values for each assay can be found in the **Appendix** (**Table A.1**) for individual studies. Following removal of these assays, samples with more than 30% missing values were removed where these occurred. Missing values for the remaining assays were replaced by the minimum or maximum analyte level for measurements below or above the limit of quantitation, respectively. Analyte values missing due to low sample volume were replaced by the mean concentration for that analyte. Missing demographic or lifestyle variable values were also replaced by the mean (continuous) or most frequent (categorical) value when they were used in the analysis. Analyte values were $\log_2$-transformed to stabilize variance and improve normality. Multivariate outliers were assessed based on robust measures of the Mahalanobis distance (see **Section 2.3.2**). Logistic, linear, and robust regressions and other analyses (see **Section 2.3**) were then carried out as described in individual chapters.

2.2 Affymetrix GeneChip® DNA microarrays

Affymetrix GeneChip® single-channel DNA microarrays were used in this work to study gene expression by measuring levels of messenger ribonucleic acid (mRNA). Single strands of deoxyribonucleic acid (DNA) are built *in situ* on a quartz wafer by photolithography with combinatorial chemistry (*286*). Unique targets are chosen from the 3' end of the transcript using various public databases (*287*). These probes are 25 bases in length, built and arranged such that each 8-11 μm square on the chip is composed of a specific type of probe (*288*). Features consist of either complementary perfect match (PM) probes or mismatch (MM) probes, illustrated in **Figure 2.2.** These differ from PM probes by one base at the centre base position and are designed to evaluate non-specific hybridization and background signal. Each sequence is interrogated with 11 probe pairs making up a probe set. The Affymetrix HG-U133A and HG-U133 Plus 2.0 chips used in this thesis measure 47,400 transcripts (from 38,500 genes) and 18,400 transcripts (from 14,500 genes), respectively (*288*).

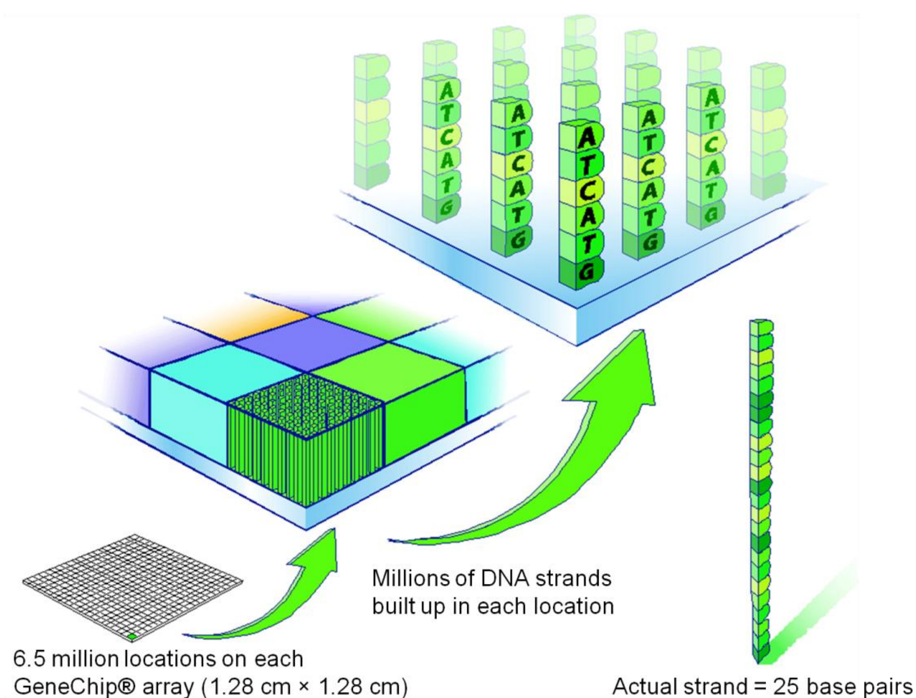


Figure 2.2. Illustration of a single Affymetrix GeneChip® feature. *Image courtesy of Affymetrix; reprinted with permission.*

RNA samples are extracted and purified from the tissue of interest, and the quantity and quality of the extracted RNA are assessed. RNA samples then undergo reverse transcription to yield complementary DNA (cDNA) followed by transcription, amplification, and fragmentation to produce biotin-labelled complementary RNA (cRNA) fragments that are hybridized to arrays in batches (*214*). Following hybridization, staining takes place with a fluorescent molecule that binds to the biotin. They are then scanned with a laser and the fluorescence intensity of the spots is quantified. A summary of this process is shown in **Figure 2.3**. Data processing is then needed for background correction, normalization, summarization, and logarithmic transformation prior to further analysis. More details on this are found in **Section 2.2.4** below.

2.2.1 Advantages

Although gene expression does not interrogate proteins directly, the high-throughput technology used to measure mRNA expression is technically simpler and more advanced than technology used to quantify protein abundance (*289*). The number of genes investigated in the DNA microarray is over 100 times the number of proteins measured with multiplex immunoassay in this thesis. Statistical methods and tools to examine quality and analyse microarray data have been developed, evaluated, and are widely available. Microarray experiments are numerous and public repositories of data have been created to store them. Measuring mRNA levels has proved useful for diagnosis and classification of cancers and other applications.

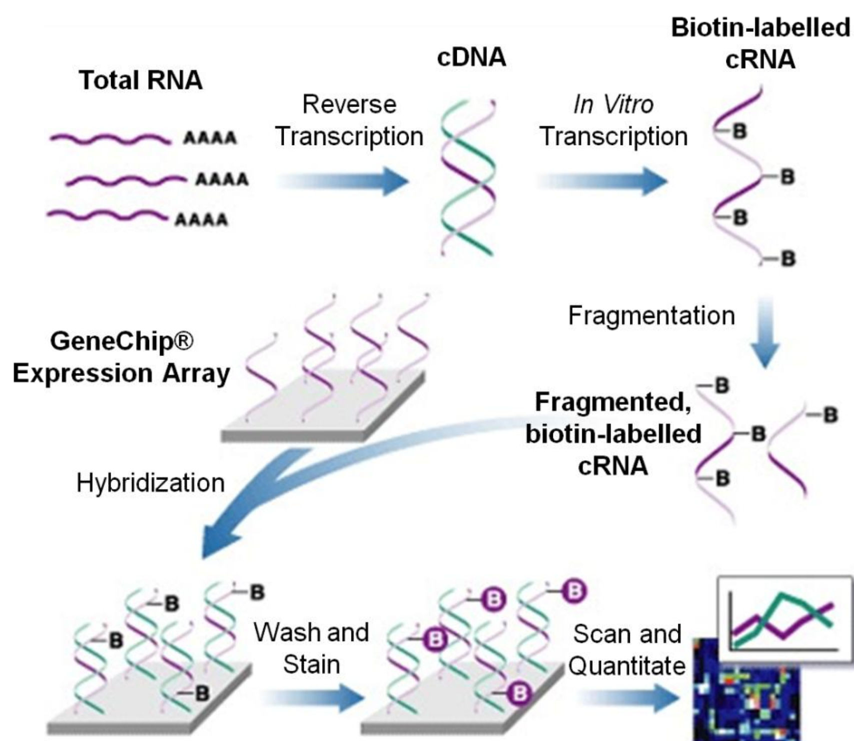


Figure 2.3. **Summary of the steps performed during an Affymetrix GeneChip® DNA microarray experiment.** *Image courtesy of Affymetrix; reprinted with permission.*

2.2.2 Disadvantages

Although proteins are translated from mRNA, a number of post-transcriptional processes may interfere with translation. This may in part cause poor correlation between levels of mRNA and protein, hindering interpretation of findings. *Schwanhäusser et al (2011)* (*290*) found that mRNA expression explained approximately 40% of the variability in the levels of proteins for 5000 genes in mammalian cells. High experimental noise occurs in microarray experiments due to sample preparation and hybridization processes.

2.2.3 Clinical samples

Five microarray datasets of schizophrenia and control prefrontal cortex (Brodmann area (BA) 9, 10, 46) were obtained from public repositories. All used Affymetrix (Santa Clara, CA, USA) HG-U133A or HG-U133 Plus 2.0 GeneChips®. Specific information for each cohort was obtained from the data repository or from articles published on the individual datasets [*Maycox et al (2009)* (*291*); *Narayan et al (2008)* (*198*)*; Glatt et al (2005)* (*214*)*; Iwamoto, Bundo, and Kato (2005)* (*292*)*; Torrey et al (2000)* (*293*)*;* the Stanley Medical Research Institute Online Genomics Database *(www.stanleygenomics.org, accessed December 2014);* and the Harvard Brain Tissue Resource Centre (Mclean66 collection, *national_databank.mclean.harvard.edu, accessed December 2014)]*. Further information is found in **Chapter 6**. Differences between groups for these variables were evaluated using

Fisher's exact test for categorical variables and a *t*-test for continuous variables, with $p<0.05$ considered significant.

2.2.4 Microarray data

2.2.4.1 Array quality

Array quality can be assessed in a variety of ways, including measurement of RNA degradation; average, minimum, and maximum background levels; percentage of genes present; scale factors; 3'/5' ratios for beta-actin and glyceraldehyde 3-phosphate dehydrogenase (GAPDH); and spike-in hybridization controls; and inspection of Relative Log Expression (RLE) and Normalized Unscaled Standard Error (NUSE) plots and spatial distribution of feature intensities. These array quality assessments were carried out here by using Bioconductor (v3.0) (*294*) arrayQualityMetrics (*295*), affy (*296*), and simpleaffy (*297*) packages with the statistical programming software R (v3.1.2) (*285*). Poor quality arrays were assigned lower weights in downstream analyses, as described in **Section 2.2.4.3** below.

2.2.4.2 Pre-processing

After array quality assessments, pre-processing reduces microarray data from the probe to the gene level. Robust multi-array average (RMA) is a widely used microarray pre-processing procedure and performs background correction, normalization, and summarization using only PM probes. In this thesis, the methodologically similar frozen RMA (fRMA) was used to pre-process microarray data (*298*). RMA and fRMA use an exponential-normal convolution model for background correction. This adjusts the signal for sources of noise in each array by modelling the observed PM intensity as $X = B + S$, where S (signal) follows an exponential distribution and B (background) follows a normal distribution truncated at zero (*299*). Background correction is performed separately for each array. Following this step, quantile normalization is used to give each array the same empirical distribution of intensities, removing technical variation and making arrays comparable. Quantile normalization requires a reference distribution, which in fRMA is provided by a "frozen" database of samples from public repositories (*298*). Probe intensities are summarized into measures of gene expression using median polish. Pre-processing of microarray data with fRMA was carried out using the Bioconductor fRMA package (*298*).

2.2.4.3 Analysis

Linear models were used to find differentially expressed genes using the limma package in Bioconductor (*300*, *301*). With the approach in limma, estimated residual sample variances for each gene are shrunk towards a pooled estimate (*301*). A moderated *t*-statistic can then

be used for inference (see **Section 2.3.4** for more on linear regression), thereby moderating standard errors for regression coefficients and increasing inference stability.

Array quality weights were used in the analysis to down-weight observations with lower quality arrays. For this, variance factors are modelled from the data, array weights are determined, and these are used to calculate a weighted least squares estimate of regression coefficients for each gene (*302*). The arrayWeight function in the R Bioconductor limma package was used to find array weights (*300*, *302*) and these were checked to ensure that low quality arrays (determined from findings in **Section 2.2.4.1**) received low weights.

2.3 Statistics

All data pre-processing, analysis, and plotting was carried out in this work using the statistical programming software R (v3.1.2) (*285*) and Bioconductor (v3.0) (*294*).This section describes most of the statistical procedures carried out in this work. Other procedures used are described in individual chapters. Procedures specific for multiplex immunoassay or microarray data are described above in **Sections 2.1.4 and 2.2.4,** respectively**.**

2.3.1 ComBat

ComBat is a method to eliminate batch effects across experiments (*303*) and was used in **Chapter 6** for microarray data and in **Chapters 3-4** for multiplex immunoassay data. It works by borrowing information across genes or proteins to calculate and adjust for additive and multiplicative batch effects. The adjusted data were then used in subsequent analyses. ComBat is implemented using the sva package in R (*304*). Principal component analysis (PCA) plots (**Section 2.3.3** below) and boxplots of individual analytes were used to assess unadjusted and batch-adjusted data.

2.3.2 Mahalanobis distance

Mahalanobis distance was used to detect multivariate outliers in **Chapters 3-5**. The Mahalanobis distance for each sample $\mathbf{x}_i$ is defined as $d_i = \sqrt{(\mathbf{x}_i - \boldsymbol{T})^T \boldsymbol{C}^{-1} (\mathbf{x}_i - \boldsymbol{T})}$, where $\boldsymbol{T}$ is a measure of location and $\boldsymbol{C}$ is an estimate of the covariance matrix (*305*). The sample mean and covariance are sensitive to outlying points, and therefore robust estimates of these were used (*306*, *307*). The classic and robust squared Mahalanobis distances have χ_p^2 and asymptotic χ_p^2 distributions for multivariate normal data, respectively (where p is the number of variables). In **Chapters 3** and **4**, the *Maronna and Zamar (2002)* (*306*) reweighted orthogonalized Gnanadesikan-Kettenring (OGK) robust estimates of location and dispersion were used to compute robust Mahalanobis distances, as implemented in covRob of the

robust package in R (*308*). In **Chapter 5**, the *Filzmoser, Maronna, and Werner (2006)* (*305*) Sign procedure, designed for data with more dimensions than the number of observations, was used as implemented in sign1 of the mvoutlier package in R (*309*).

2.3.3 Principal component analysis (PCA)

Principal component analysis (PCA) was used in **Chapters 3-6** to reduce the dimensionality of the data in order to facilitate visualization of patterns, including plate/batch effects, in the data. PCA converts correlated variables into sets of uncorrelated linear combinations of these variables (called principal components, PCs) that capture as much information as possible about the data. Geometrically, for the first PC, PCA finds the vector (called the loading vector) that is the direction along which the greatest variation occurs in the data in the space of the original variables. Scores are the projections of the data points onto this vector. The second PC is found in the same way with the constraint that the loading vector for the second PC is orthogonal to the loading vector for the first PC (the scores are uncorrelated with those of the first PC). Subsequent PCs are found such that the mth PC loading vector is the direction of maximum variance, subject to being orthogonal to all previous loading vectors. PCA assumes that large variances have important structure, that PCs are orthogonal, and linearity (*310*). The PCA function in the R package FactoMineR was used for PCA analysis (*311*).

2.3.4 Linear regression

Linear regression models a linear relationship between a dependent variable and one or more explanatory variables in **Chapters 3-6** of this thesis. This model is represented as:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_p x_{ip} + \varepsilon_i\,, \qquad i = 1, \dots, n$$

where y_i is the dependent variable for observation i; $x_{i1}, \dots, x_{ip}$ are the p independent variables for observation i; $\beta_1, \dots, \beta_p$ are regression coefficients; n is the number of observations; and ε_i is an unobserved random variable representing noise added to the linear relationship. An intercept is included by setting $x_{i1} = 1$ for $i = 1, \dots, n$ and is the predicted value of the dependent variable when the values of all independent variables are zero. Coefficients for independent variables can be interpreted as the predicted change in the dependent variable for a one unit change in the independent variable, holding all other independent variables constant (i.e., the value of the partial derivative of the dependent variable with respect to that independent variable). When multicollinearity is present among the independent variables, the regression coefficient of each independent variable represents its additional contribution on the dependent variable after it has been adjusted for the remaining independent variables (*312*).

Categorical independent variables can be included in the regression by recoding to a set of new variables, with the number of new variables equal to one less than the number of categories. Dummy coding was used in **Chapters 3-6**, in which one category is chosen as the reference and indicator variables represent remaining categories. Each new variable is coded as 1 to indicate membership in a particular category and 0 otherwise (*312*). Coefficients for dummy coded categorical variables can be interpreted as differential intercept coefficients, whereby the intercept for the regression is shifted between categories. In **Chapter 3**, contrasts were set manually to compare categories.

Ordinary least squares (OLS) estimates of β, found by minimising the sum of squares of the residuals, have the smallest variance amongst unbiased linear estimates according to the Gauss-Markov theorem (assuming linearity, constant error variance, independence, and lack of linear dependence in the independent variables). The lm function in the stats package in R (*285*) was used to carry out OLS linear regression and hypothesis testing for regression coefficients.

2.3.4.1 Assumptions of OLS

The following assumptions should be met in order to guarantee the validity of OLS estimates and other desirable properties described below (*313–315*):

1. *Correct specification:* the model should be correctly specified with linear parameters.
2. *Measurement error:* independent variables are measured without error.
3. *The expected value of errors in the regression is 0, conditional on the independent variables* $[E(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \mathbf{0}]$*:* on average the model is exactly linear and errors are not correlated with independent variables.
4. *Spherical errors* $[\mathrm{Var}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \sigma^2\boldsymbol{I_n}]$*:*
    a. *Homoskedasticity* $\left[E\left(\varepsilon_i^2|\boldsymbol{X}\right) = \sigma^2, i = 1, \dots, n\right]$: constant error variances. When heteroskedasticity is present, OLS estimates will not be efficient and inaccurate standard error calculations will result (*315*).
    b. *No serial correlation* $\left[E\left(\varepsilon_i\varepsilon_j|\boldsymbol{X}\right) = 0, i, j = 1, \dots, n, i \neq j\right]$: error terms of observations are uncorrelated.
5. *No linear dependence in the independent variables.* Strong collinearity with other independent variables will cause the coefficient to be highly unstable and inflate its variance.
6. *Normality of errors* $[\boldsymbol{\varepsilon}|\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I_n})]$: this assumption is required for hypothesis testing and construction of confidence intervals. In large samples the levels of hypothesis tests and confidence intervals are approximately correct even with non-

normal errors (*315*). However, particularly with heavy-tailed error distributions, OLS will be inefficient (*315*).

In this thesis, these assumptions were verified by: examining residual regression plots to check for normality, heteroskedasticity, and examine influential points; checking for duplicate samples; and computing correlations between independent variables. As mentioned previously, analyte levels were log_2-transformed to stabilize variance and improve normality.

2.3.5 Robust regression

Unusual data can heavily influence OLS. Regression outliers are observations with unusual values conditional on the independent variable. These were identified in this thesis as Studentized regression residuals with Bonferroni-corrected *p*-values < 0.05 (used to control the probability of making one or more false discoveries), calculated using the outlierTest function in the car package in R (*316*). When outliers occur at high leverage points, they have a substantial influence on the regression coefficients (*315*). Robust regression was performed as an alternative to OLS in this case in **Chapters 3, 5,** and **6**. MM-estimators for regression were used, implemented in lmrob (robustbase package) in R (*317*), as they have a 50% breakdown point and 95% efficiency relative to OLS (*318*, *319*). The following procedure is used to find an MM-estimator of regression:

1. Regression coefficients are found first using a high breakdown, consistent, but not necessarily efficient estimator. In this thesis, S-estimators are used, which find regression coefficients that minimise a robust estimate of the scale of the residuals. Regression residuals from the S-estimator are calculated.
2. These residuals are used to compute a robust estimate of scale with a breakdown point of 50%.
3. Finally, regression coefficients are computed using an efficient, robust estimator (M-estimator) with the computed estimate of scale from the second step.

MM-estimates of regression coefficients are asymptotically normal, enabling hypothesis testing, conducted using lmrob (robustbase package) in R (*317*).

2.3.6 Generalized least squares (GLS)

GLS was used in **Chapter 4** to account for heteroskedasticity (non-constant error variance). Regression coefficients are found by solving a weighted ordinary least squares problem, as implemented using the gls function in the nlme package in R (*320*). Hypothesis tests were conducted using a likelihood ratio test (see **Section 2.3.7** below).

2.3.7 Logistic regression

Logistic regression models the relationship between the probability of an event $[\pi(X) = \Pr(Y = 1|X)]$ and the independent variables in **Chapter 4**. The logistic function is (*321*, *322*):

$$\pi(\mathbf{x}_i) = \frac{e^{\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3}+\cdots+\beta_p x_{ip}}}{1 + e^{\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3}+\cdots+\beta_p x_{ip}}}, \qquad i = 1, \ldots, n$$

where observations are indexed with subscript i; n is the number of observations; and $\beta_1, \ldots, \beta_p$ are regression coefficients for p explanatory variables. This model can be manipulated to yield a function that is linear in the independent variables, with the left-hand side termed the logit:

$$\log\left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}\right) = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_p x_{ip}, \qquad i = 1, \ldots, n$$

Parameters for this model are found using maximum likelihood estimation, a general method for estimating model parameters. Regression coefficients are interpreted as the change in log-odds associated with a one unit change in the independent variable, holding constant/adjusting for all other independent variables in the model. Exponentiating the coefficient gives the odds ratio (OR) associated with a one unit change in the independent variable:

$$OR = e^{\beta_j} = \frac{odds\ when\ x_j = h + 1}{odds\ when\ x_j = h}, \qquad j = 1, \ldots, p$$

for p explanatory variables. A likelihood ratio test was conducted to compare the fit of the full model with the reduced model (*322*). Logistic regression and hypothesis testing was carried out using the glm function in the stats package of R (*285*).

2.3.7.1 Assumptions of logistic regression

Violation of logistic regression assumptions can have consequences for bias, efficiency, and statistical inference of estimates (*321*). The following assumptions are made:

1. *Correct specification:* The logistic model should be correctly specified, with the binomial distribution appropriate for the response variable. The relationship described by the model should be linear in the logit (*321*).
2. *No linear dependence in the independent variables.* Strong collinearity with other independent variables will cause the coefficient to be highly unstable and inflate its variance (*321*).

3. *No overdispersion or underdispersion.* These occur when the observed variance is smaller or greater than the expected variance, respectively (*321*). However, in Bernoulli (binary response) data used in this thesis, there is no overdispersion or underdispersion.
4. *Independence of errors* (*323*).

As in linear regression, regression outliers at high leverage points can heavily influence model fit. In this thesis, assumptions were verified by examining residual plots for outliers, checking for duplicate samples, and computing correlations between independent variables.

2.3.8 Interactions

Interactions between variables occur in regression analyses when the relationship between the dependent variable and an independent variable is modified by one or more independent variables. Here, the relationships between sex, schizophrenia or MDD diagnosis, and analyte levels were explored in linear and logistic regression. In linear regression, the model evaluated (**Chapters 5 and 6**) was:

$$y_{ik} = \beta_{0k} + \beta_{1k}[sex_i] + \beta_{2k}[diagnosis_i] + \beta_{3k}[sex_i \times diagnosis_i] + \beta_{4k}x_{i4} + \cdots + \beta_{pk}x_{ip} + \varepsilon_{ik},$$
$$k = 1, \dots, m, \qquad i = 1, \dots, n$$

where y_{ik} is the log₂-transformed concentration of analyte k for observation i; sex_i, $diagnosis_i$, and $sex_i \times diagnosis_i$ represent these variables; $x_{i4}, \dots, x_{ip}$ are $p - 3$ additional explanatory variable vectors; $\beta_{0k}, \dots, \beta_{pk}$ are regression coefficients; ε_{ik} is an unobserved random variable representing noise; m is the number of analytes; p is the number of explanatory variables; and n is the number of observations. **Sections 2.3.4 and 2.3.5** above provide more details on linear regression (OLS and robust).

Main effects of sex and diagnosis (β_{1k} and β_{2k}, respectively) are the predicted differences in $\log_2$-transformed analyte levels between males and females (β_{1k}) and between schizophrenia patients and controls (β_{2k}), adjusted for/holding constant other variables in the model. Their interaction, β_{3k}, is the predicted change in the relationship between the $\log_2$-transformed analyte level and diagnosis in females compared to males. Since analyte values are $\log_2$-transformed, testing whether the interaction coefficient $\beta_{3k} = 0$ is equivalent to testing whether $g_{3k} = 2^{\beta_{3k}} - 1 = 0$. This can be interpreted as assessing whether the proportional impact of schizophrenia diagnosis on the raw, untransformed serum concentration of analyte k is the same in males and females (adjusted for/holding constant other variables in the model) (*324*):

$$\frac{y_{raw,k(male,SCZ)}}{y_{raw,k(male,NC)}} \bigg/ \frac{y_{raw,k(female,SCZ)}}{y_{raw,k(female,NC)}} = 1, \quad k = 1, \dots, m$$

where $y_{raw,k}$ is the untransformed concentration of analyte k, SCZ is schizophrenia, and NC is control. When this was significant, differences in analyte levels between schizophrenia patients and controls were investigated separately for males and females (i.e., separate estimation of β_{2k} $\left(g_{2k} = 2^{\beta_{2k}} - 1\right)$ for males and females):

$$g_{2k,male} = \frac{y_{raw,k(male,SCZ)}}{y_{raw,k(male,NC)}} - 1, \qquad g_{2k,female} = \frac{y_{raw,k(female,SCZ)}}{y_{raw,k(female,NC)}} - 1, \qquad k = 1, \dots, m$$

Example plots of potential main effects and interactions between sex and schizophrenia diagnosis in linear regression are in **Figure 2.4.**

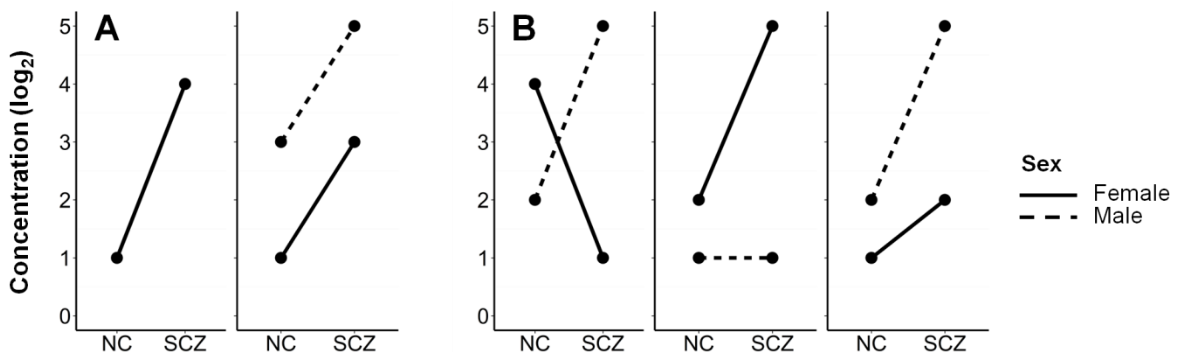


Figure 2.4. **Illustration of main effects and interactions. (A)** shows a main effect of diagnosis on log_2-transformed concentration of analyte $\boldsymbol{k}$ on the left, and main effects of both diagnosis and sex on the right. **(B)** shows examples of possible interactions between sex and diagnosis. **Abbreviations:** NC (control); SCZ (schizophrenia).

In logistic regression, the model evaluated (**Chapter 4)** was:

$$\log\left(\frac{\pi(\mathbf{x}_{ik})}{1-\pi(\mathbf{x}_{ik})}\right) = \beta_{0k} + \beta_{1k}[sex_i] + \beta_{2k}[y_{ik}] + \beta_{3k}[sex_i \times y_{ik}] + \beta_{4k}x_{i4} + \cdots + \beta_{pk}x_{ip},$$
$$k = 1, \dots, m, \qquad i = 1, \dots, n$$

Section 2.3.7 above provides more details on logistic regression. Testing the hypothesis $\beta_{3k} = 0$ assesses whether the relationship between the log odds of MDD and the log_2-transformed concentration of analyte k differs between males and females (again, adjusted for/holding constant other variables in the model). When this occurred, inference and estimation on β_{2k} was conducted for males and females separately:

$$OR_{k,male} = e^{\beta_{2k,male}} = \frac{odds\ of\ MDD\ in\ males\ when\ y_{raw,k(male)} = 2h}{odds\ of\ MDD\ in\ males\ when\ y_{raw,k(male)} = h},$$

$$OR_{k,female} = e^{\beta_{2k,female}} = \frac{odds\ of\ MDD\ in\ females\ when\ y_{raw,k(female)} = 2h}{odds\ of\ MDD\ in\ females\ when\ y_{raw,k(female)} = h}, \quad k = 1, \ldots, m$$

Since analytes are $\log_2$-transformed, the exponentiated regression coefficient gives the odds ratio associated with a two-fold increase in the untransformed analyte concentration (again, adjusted for/holding constant other variables in the model).

Interactions were also used to assess heterogeneity in effects between cohorts in **Chapters 5** and **6**, where multiple studies were pooled together. This involved testing for an interaction between the cohort variable and the variable of interest (diagnosis or sex-diagnosis interaction term).

2.3.9 Model selection

In order to avoid overfitting and to clarify interpretation, variable subset selection can be used in regression (*325*). Analysis of variance (ANOVA) compares nested linear models in **Chapter 3** and tests the null hypothesis that new variables do not add explanatory value over the reduced model. This was carried out using the aov or anova functions in the stats package in R (*285*). ANOVA for robust regression was also used, as implemented in the robustbase package in R (*317*).

Stepwise variable selection was used in **Chapters 3-5** for model building. Forward-stepwise selection was used, which sequentially adds independent variables to the model to improve the fit until a stopping criterion is reached, as well as mixed (or backward- and forward-) stepwise selection, which considers removing or adding a variable at each step. Bayesian information criterion (BIC) was used to select the most parsimonious model (*325*), which provides an indirect estimate of the test error of the model by adjusting the training error to account for the number of parameters in the model. BIC generally penalizes larger models more heavily than other criteria and is asymptotically consistent, selecting the correct model as $n \to \infty$ (*312*). The step function in the stats package of R (*285*) was used to perform stepwise regression.

2.3.10 False discovery rate (FDR)

Performing multiple statistical hypothesis tests simultaneously increases the chances of incorrectly rejecting a null hypothesis. In the microarray and proteomic experiments performed in this work, hundreds or thousands of hypothesis tests are performed. False discovery rate (FDR) procedures were therefore used in **Chapters 3-6** to control the expected proportion of false positive findings. The method of *Benjamini and Hochberg (1995)* (*326*) assumes independent tests and defines the false discovery rate:

$$FDR = E\left(\frac{V}{R}|R > 0\right)\Pr(R > 0)$$

where V and R are given in **Table 2.1** below, which describes the outcomes of testing m hypotheses:

Table 2.1. Results of performing m hypothesis tests: correct and incorrect rejections or failures to reject the null hypothesis.

	Do not reject the null hypothesis	Reject the null hypothesis	Total
Null hypothesis true	U	V	m_0
Alternative hypothesis true	T	S	m_1
	W	R	m

The Benjamini and Hochberg FDR controlling procedure is performed by first ranking the m obtained *p*-values in order of significance. Then k is the largest i for which

$$P_{(i)} \leq \frac{i}{m}\alpha$$

where i is the index of the ranked *p*-value and α is the level at which FDR is controlled. Then all hypotheses $i = 1, \ldots, k$ are rejected and $E(V/R) \leq \alpha$. A FDR analogue of the *p*-value, called the *q*-value, is the expected proportion of false positives among the list of results up to that finding, or the minimum FDR that can be attained where the finding will be called significant. These are calculated from the ordered *p*-values:

$$q_{(i)} = min\left\{q_{i+1}, \frac{m}{i}P_{(i)}\right\}, \quad i = 1, \ldots, (m-1)$$
$$q_{(m)} = P_{(m)}$$

The p.adjust function in the stats package in R was used to compute *q*-values in this work (*285*).

2.3.11 Random forests

Random forests were used in **Chapter 3** to classify observations using analyte concentrations and other demographic variables. This ensemble method constructs many classification trees and classifies new observations based on the majority decision (majority vote) of these trees (*327*). Each tree in the forest is grown from n random cases sampled with replacement (from n observations) by recursively splitting observations until the terminal node contains data purely from one category (*328*, *329*). The maximum decrease in the Gini impurity criterion, a measure of node homogeneity, determines the best way to split the data

at each node from $\sqrt{p}$ (of p total) randomly selected variables (*328*, *329*). The Gini impurity criterion is (*312*):

$$I_G = 1 - \sum_{i=1}^{m} f_i^2$$

where f_i is the fraction of observations with label i in the set, of a total of m possible labels. Greater node homogeneity occurs with smaller values of I_G, which reaches zero when the node contains observations from only one category. **Figure 2.5** shows a simulated example of a single decision tree for classifying an observation as either disease or control based on $\log_2$-transformed analyte concentrations. Random forests were composed of 5000 trees and built using the randomForest function in the R package of the same name (*329*). Variable importance was assessed by adding the decreases in Gini impurity resulting from a split on a particular variable and averaging over the number of trees (*328*).

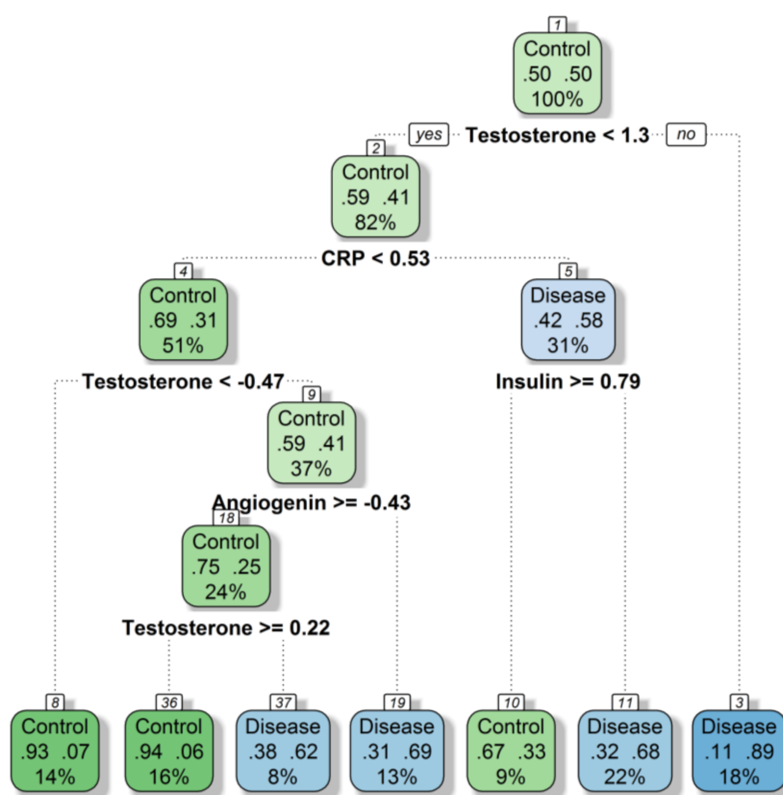


Figure 2.5. Simulated decision tree for classifying an observation as either disease or control based on $\log_2$-transformed analyte concentrations. For each node, the top line shows the enriched class; the middle line shows the proportion of each class; and the bottom line shows the percentage of observations. The decision boundary is shown below each node. Only part of the decision tree is shown (terminal nodes are not shown).

2.3.12 Classification performance

In order to assess the performance of the MDD classifications in **Chapter 4**, a repeated ten-fold cross-validation procedure was performed, followed by plotting the receiver operating

characteristic (ROC) curve and measuring the area under the ROC curve (AUC). For the ten-fold cross-validation, the data was randomly divided into ten partitions, stratified by sex and MDD diagnosis using the sample function in the base package in R. Nine of these partitions were used to create a logistic model that was used to estimate the probability of MDD diagnosis for the observations in the last partition. This was then repeated for each of the remaining partitions. The entire process of randomly dividing into stratified partitions, building the model using nine partitions, and evaluating in the last was done 50 times. This provides a better estimate of the complete cross-validation (or leave-10-out cross-validation), the average of all possible different splits into ten partitions (*330*). The average ROC curve can be generated by first merging all test sets into one large set (*331*). The true and false positive rates are then found and plotted for each classification threshold. A classifier with a ROC curve along the line $y = x$ represents the performance expected by randomly guessing a class, with curves above this line representing better classifier performances. The AUC of the ROC curve is a single measure of classifier performance and is the probability that a randomly chosen subject with MDD will be ranked higher than a randomly chosen control subject (*331*, *332*). Examples of possible ROC curves and AUCs for classifiers are shown in **Figure 2.6.** In this thesis, ROC curves were generated and AUCs were calculated using the ROCR package in R (*333*).

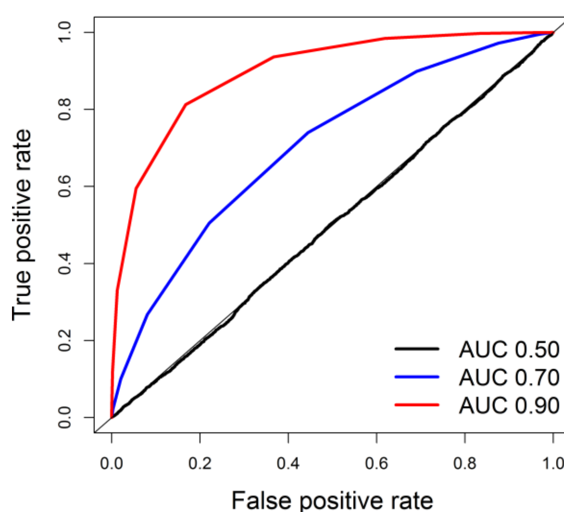


Figure 2.6. Examples of ROC curves to assess classification performance and their AUCs.

2.4 Gene Ontology (GO) terms

The Gene Ontology (GO) Consortium provides a dynamic, controlled vocabulary for annotating genes based on three independent ontologies: cellular components, molecular functions, and biological processes (*334*). **Chapters 3-6** investigated the biological

processes of genes and proteins found to be involved in sex and mental disorders using GO terms and manual searches of the literature. UniProtKB accession (proteins) and Affymetrix 3' IVT array identifiers (genes), were used in the Uniprot.ws (*335*) and biomaRt (*336*, *337*) packages in Bioconductor, respectively, to find biological process GO terms associated with lists of genes or proteins. These were mapped to ancestor terms using the GOSim package in Bioconductor (*338*). In **Chapters 3** and **6**, summaries of GO terms for significant genes or proteins were prepared, with selected terms based on the number and relevance of annotated proteins. The European Bioinformatics Institute (EBI) browser for GO annotations, QuickGO (http://www.ebi.ac.uk/QuickGO/, *accessed March 2015*) and the UniProt website (*www.uniprot.org; accessed March 2015)* were also used to investigate GO terms for individual genes or proteins.

2.4.1 GO term enrichment

GO biological process term enrichment of significant genes was examined in **Chapter 6.** Affymetrix 3' IVT array identifiers were mapped to Entrez Gene identifiers using biomaRt (*336*, *337*) and overrepresented biological process (BP) terms were evaluated using GOstats in Bioconductor (*339*). GOstats uses a hypergeometric test to evaluate enrichment of terms. However, the hierarchical structure of GO ontologies often means many directly related GO terms are identified as being enriched. For this reason, a conditional test was used in which genes with significantly enriched terms are removed when testing the parent term (*339*). Here, at least ten annotated genes for the term were required and $p<0.001$ was considered statistically significant.

Chapter 3 Variations in serum molecular concentrations associated with sex and female hormonal status

3.1 Introduction

Sex and gender differences are an important component of the heterogeneity of mental disorders and can include differences in prevalence, incidence, age of onset, symptoms, course of the disorder, and treatment response (*17*, *154*, *157*, *236*, *340*). Some of these may be explained by sex-dependent regulation of biological pathways. Endocrine, nervous, immune, metabolic, and cardiovascular function all show substantial variation between males and females and are important in mental disorders (see **Chapter 1**). However, many of these processes and the changes that may accompany variations in the levels of sex hormones after menopause, during oral contraceptive (OC) use, and within the menstrual cycle are poorly understood.

Important insights into sexually dimorphic disease mechanisms have been gained by studying sex differences in control populations. Research into sex differences in immune, metabolic, and cardiovascular function has benefitted understanding of the higher prevalence of most autoimmune diseases in females, better prognosis for cardiovascular diseases in premenopausal females, sex differences in pharmacological response, and others (see **Chapter 1**). The study of nervous system development and function in males and females has also lead to a better understanding of possible mechanisms for sex differences in mental disorders, though many other potentially important sexually dimorphic biological processes remain to be elucidated.

Investigations of sex differences at the molecular level carried out so far have largely been limited to the study of only a few molecules and mechanisms. A previous study by *Ramsey et al. (2012)* (*341*) evaluated sex differences in 174 serum molecules in a meta-analysis of 9 independent cohorts comprising 392 subjects free of mental disorders and other chronic conditions (196 males/196 females). They found 40 analytes elevated in females and 37 elevated in males. Clusters of molecules enriched in fatty acid oxidation/hormone regulation, immune cell growth and activation, and cell death processes were elevated in females, while clusters of molecules enriched in immune cell chemotaxis and other indistinct functions were elevated in males. However, demographic and lifestyle information was limited and hormonal status of females (phase of menstrual cycle, menopausal status, use of OCs, etc.) was unavailable. Furthermore, although the *Ramsey et al (2012)* (*341*) study found sex

differences in a large proportion of serum molecules, the potential consequence of these differences for biomarkers studies was not evaluated. Although many studies now record and some also match for sex, female hormonal status is typically not assessed. This includes recent biomarker studies of schizophrenia and major depressive disorder (MDD) (*129*, *278*). The result of ignoring sex and female hormonal status in these and other studies has not been assessed.

3.1.1 Aims and objectives

The work presented in this chapter had the following aims and objectives:

1.) To investigate differences in biological processes between males and females by determining serum analyte concentrations that vary with sex and female hormonal status (OC use, phase of menstrual cycle, and menopausal status). This was done in a cohort of subjects free of mental disorders using a Human DiscoveryMAP® multiplex immunoassay panel of 243 molecules. Agreement with the *Ramsey et al (2012)* (*341*) meta-analysis was evaluated and the potential relevance of the results to schizophrenia and MDD was assessed by comparing them to reported peripheral (serum/plasma) markers of these disorders.

2.) To evaluate the possible effect of ignoring sex and female hormonal status on the number of false positives obtained in biomarker studies comprising these analytes. This was assessed through simulations of group imbalances in these variables.

3.) To construct and evaluate algorithms for classifying sex and female hormonal status of samples. Classification tools were constructed using both the full set of analytes and demographic, lifestyle, and health variables available in the current study and also using a reduced set that was available in previous studies. Performance of the classification tool was evaluated on subjects with current and/or past depressive and anxiety disorder episodes. The goal of this was to assess whether sex and/or female hormonal status of samples from prior investigations could be evaluated where this had not been recorded.

In order to accomplish these aims and objectives, samples from the Netherlands Study of Depression and Anxiety (NESDA) were used. NESDA is a well characterized and ongoing 8-year longitudinal, multi-site naturalistic cohort study with participants recruited from the general population, general practices, and mental health organizations in the Netherlands (*342*). A total of 347 individuals free of mental disorders enrolled in the NESDA study were included in the first two parts of the study and another 1,329 suffering from current and/or

past depressive and anxiety disorders episodes were also included in the last part of the study.

3.2 Methods

3.2.1 Clinical Samples

Serum samples from NESDA were collected from municipalities in the Netherlands. Recruitment of 2,981 participants aged 18-65 years took place between 2004-2007 from the community (19%), general practice (54%), and secondary mental health care (27%). The ethical committees of all participating universities approved the protocols of the study and informed written consent was given by all participants. Local ethical approval for use of human serum samples was granted by the Cambridgeshire Local Research Ethics Committee.

Control subjects had neither a current nor a lifetime diagnosis of assessed psychiatric disorders according to the Composite Interview Diagnostic Instrument (CIDI) and did not develop any such disorder by the second year follow-up assessment. In women, use of OCs and sex hormones (World Health Organization Anatomical Therapeutic Chemical (WHO-ATC) classification code G03; yes/no) was assessed by self-report. OCs were combined (estrogen and progestogen) or progestogen only with varying doses of estrogen, different types of progestogen, and were monophasic or triphasic. Classification of menstrual cycle phases in women [follicular (0-13 days) and luteal (14-32 days or more)], postmenopausal status (yes/no), hysterectomy (yes/no), and pregnancy or breastfeeding status (yes/no) were also self-reported. Women who were pregnant or breastfeeding, using hormones not classified as OCs, had ungergone a hysterectomy, or with missing information on hormonal status were excluded from further analysis in this study. Other demographic, lifestyle, and health information recorded included ancestry (North European/other); education (years); systolic and diastolic blood pressure (mm Hg); physical activity (measured in 1000 metabolic equivalent (MET)-minutes per week, assessed with the International Physical Activity Questionnaire (*343*)); self-reported presence of treated chronic somatic disease (yes/no, further details below); use of lipid modifying agents (ATC code C10; yes/no); use of anti-inflammatory drugs (ATC codes H02, R03BA, R03AK, D07, M01A, M01B, A07EB or A07EC; yes/no); use of antihypertensive medication (ATC codes C02, C03, C07, C08, or C09; yes/no); collection area (Amsterdam, Leiden, and Groningen); body mass index (BMI; kg/m^2); waist circumference (centimetres); age (years); alcohol consumption (drinks/week); smoking status (never smoked/former smoker/regular smoker/not a regular smoker);

recreational drug use in the past month (yes/no, further details below); partner status (with/without partner); sex (male/female); and female hormonal status (see above). Chronic diseases considered were cardiovascular disease, diabetes, lung disease, osteoarthritis, rheumatic disease, cancer, ulcer, intestinal problem, liver disease, epilepsy, thyroid gland disease and others. Presence of cardiovascular disease was assessed by self-report and confirmed by use of appropriate medication, while presence of diabetes was assessed by use of anti-diabetic medication (ATC code A10) or by fasting plasma glucose level ≥7.0 mmol l^{-1}. Other chronic diseases were treated and assessed by self-report. Recreational drugs considered were cannabis, speed, cocaine, heroin, and LSD. Medication use was evaluated by self-reporting and inspection of drug containers used in the past month, from which ATC codes were recorded.

Diagnoses of depressive disorders (MDD and dysthymia) and anxiety disorders (social phobia, generalized anxiety disorder, panic disorder, and agoraphobia) were carried out during a baseline interview by specially trained research staff using the CIDI in accordance with the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV criteria (*22*). Patients with current disorders were classified as having a depressive and/or anxiety episode within the past six months. Participants completed a 30-item self-rated Inventory of Depressive Symptomatology (IDS) and 21-item self-report Beck Anxiety Inventory (BAI) (*344*, *345*). Remitted patients had a lifetime diagnosis of a depressive and/or anxiety disorder but were not currently experiencing an episode. Patients and controls were excluded who were not fluent in the Dutch language, had a primary clinical diagnosis of other psychiatric disorders not studied in NESDA: bipolar disorder, obsessive compulsive disorder, severe substance use disorder, or psychotic disorder. Further details of the NESDA study design and protocol and descriptions of all participant information recorded have been previously described (*342*).

Blood samples from NESDA were collected at the baseline assessment at approximately 0800 hours from all participants. The Human DiscoveryMAP® multiplex immunoassay platform, described in **Chapter 2,** was used to measure the serum concentrations of 243 total analytes across 1,840 available NESDA participant samples assigned to 26 plates. Remaining samples did not participate in both baseline and two-year follow-up assessments, had insufficient volume, or were otherwise unusable. Of these, three samples were removed after internal quality control checks.

3.2.2 Data pre-processing

Data analysis and pre-processing was carried out as described in **Chapter 2**. Seventy-two assays were removed with more than 30% missing values, leaving 171 for further analysis.

A list of these 171 analytes, their UniProtKB accession or PubChem compound identifiers, and the percentage of missing values in each assay can be found in the **Appendix** (**Table A.1**)**.** In addition, one sample with more than 30% missing assays was removed. Remaining missing assay values were replaced as described in **Chapter 2.** To adjust for batch effects caused by running samples on different plates, ComBat was used after $\log_2$ transforming the analyte data, as described in **Chapter 2**. Multivariate outlier assessment (**Chapter 2**) resulted in the removal of an additional four samples. Missing covariate data (MET: 5.0% missing, alcohol consumption: 0.8% missing, and recreational drug use: 0.5% missing) was replaced with the mean or most frequent value for continuous and discrete variables, respectively.

3.2.3 Sample exclusion

After data pre-processing, 1,832 samples remained for analysis. Additional samples were excluded for the following reasons: 1) the participants had not fasted when blood was withdrawn (N=56); or the participants were females who, 2) were pregnant or breastfeeding (N=10), 3) used sex hormones not classified as OCs (N=18), 4) had undergone a hysterectomy (N=61), and 5) had missing hormonal status information (N=11). This left 1,676 samples for further analysis, consisting of 347 control subjects (140 males/207 females) and 1,329 subjects with a current or remitted diagnosis of depression, anxiety, and/or dysthymia, or who were diagnosed with a mental disorder in the second year follow-up assessment (261 males/1068 females).

3.2.4 Data analysis

A linear model was used to detect sex differences in the concentrations of analytes in NESDA control subjects. Demographic and other information for males and females is summarized in **Table 3.1A**. The analysis also considered collection area (Amsterdam, Leiden, and Groningen in the Netherlands); recruitment method (community/general practice); chronic disease (yes/no); use of lipid modifying agents (yes/no); use of anti-inflammatory drugs (yes/no); use of antihypertensive medication (yes/no); ancestry (North European/other); education (years); physical activity (1000 MET-minutes per week); body mass index (BMI; kg/m^2); age (years); alcohol consumption (drinks/week); smoking status (never smoked/former smoker/regular smoker/not a regular smoker); recreational drug use (yes/no); partner status (with/without partner); and IDS score as additional covariates for this analysis of analyte data. Covariates were selected using stepwise regression with simultaneous forward selection and backward elimination using BIC selection criterion, as described in **Chapter 2**. Robust regression (MM-estimator) was used where regression outliers were present (with Bonferroni-corrected p-value < 0.05), calculated using the

robustbase package in R (*317*). The $\log_2$ ratios of serum molecular concentrations in males compared to females were found from the coefficients of the linear regression. Gene ontology (GO) biological process (BP) terms for significant proteins were found using the R Bioconductor package Uniprot.ws (*335*) and these were mapped to all ancestor terms using the *GOSim* package (*338*). A summary of GO terms for each significant protein was prepared, with GO terms selected based on relevance and number of annotated proteins. The results were compared to the *Ramsey et al (2012)* (*341*) meta-analysis findings. Furthermore, the relevance of the findings to schizophrenia and MDD were evaluated. This was done by compiling lists of reported peripheral markers (plasma or serum) of schizophrenia (*130*, *209*, *212*, *278*, *346*) and MDD (*129*, *130*) and comparing these to analytes that varied significantly with sex and hormonal status in the current study.

In order to investigate the effect of female hormonal status on concentration of serum analytes, ANOVA was used to evaluate the effect of adding these (male/follicular phase/luteal phase/postmenopausal status/OC use) to the linear model. Where robust regression was used, ANOVA used a robust Wald-type test implemented in the robustbase package in R (*317*). Contrasts were used to evaluate the levels of serum analytes in females using OCs compared to females with a menstrual cycle (females in the follicular and luteal phases of menstruation), postmenopausal compared to females with a menstrual cycle, males compared to females with a menstrual cycle, and females in the follicular compared to luteal phase of the menstrual cycle. The $\log_2$ ratios of serum molecular concentrations between hormonal statuses were assessed. Finally, the first two principal components (PCs) from a principal component analysis (PCA) of all analyte data were used to visualize patterns of variation arising with sex and female hormonal status.

The FDR analogue of the *p*-value (*q*-value) was calculated for these analyses to control the expected proportion of false positive findings due to multiple testing, described in **Chapter 2**. *Q*-values were considered significant at the 5% level.

Table 3.1. Demographic, lifestyle, and health characteristics for (A) NESDA controls (training set) and (B) NESDA test set subjects. The set of subjects in **(A)** was used to evaluate variations in serum analyte levels with sex and female hormonal status and to construct random forest classifiers to group subjects as males, oral contraceptive (OC) users, postmenopausal females, and females with a menstrual cycle. The set of subjects in **(B)** was used as a test set for the random forest classifiers built using the NESDA control cohort. Values are shown as the mean ± the standard deviation. Differences between groups were assessed using ANOVA (continuous data) or Fisher's exact test (categorical data). Variables in bold are significantly different ($p < 0.05$) between groups. **Abbreviations**: BMI (body mass index); IDS (Inventory of Depressive Symptomatology).

(A)	**Female**				**Male**
	OC user	**Follicular phase**	**Luteal phase**	**Postmenopausal**	
N	79	34	37	57	140
Age (years)	28±10	36.4±12	35.8±10	55.9±6	39.7±15
BMI (kg/m^2)	23.4±4	25.1±6	25.1±5	25.8±5	25.3±4
Waist circumference (cm)	79.5±12	82.5±13	85.4±14	88.3±12	93.1±13
North European ancestry % (Yes/No)	96/4	94/6	100/0	98/2	99/1
Area % (Amsterdam/Groningen/Leiden)	10/68/22	15/38/47	22/41/38	26/23/51	11/51/38
Frame % (General population/Primary care)	48/52	29/71	27/73	0/100	33/67
Education (years)	12.6±3	13.1±4	13.7±3	11.9±4	12.7±3
Partner % (Yes/No)	73/27	79/21	76/24	82/18	74/26
Smoking % (Current smoker/Former smoker/Never smoker)	19/30/51	32/35/32	30/19/51	23/44/33	31/37/32
Alcohol (drinks per week)	4.4±5	2.8±5	5.5±7	6.6±8	10.8±11
Recreational drug use in last month % (Yes/No)	5/95	9/91	3/97	0/100	9/91
Chronic disease % (Yes/No)	19/81	38/62	22/78	51/49	31/69
Anti-inflammatory medication % (Yes/No)	3/97	3/97	5/95	4/96	1/99
Lipid modifying agents % (Yes/No)	1/99	0/100	0/100	7/93	8/92
Antihypertensive medication % (Yes/No)	9/91	3/97	3/97	23/77	15/85
Physical activity (MET)	3822.4±2571	4616.2±4093	3079.5±2595	3626.2±2687	4154.5±3551
Systolic blood pressure (mm Hg)	127.6±15	122.5±12	124.7±14	139.5±28	144.8±18
Diastolic blood pressure (mm Hg)	77.3±10	74.4±8	76.8±11	80.7±15	81.7±12
IDS score	8.5±7	8.8±9	7.3±6	9.8±8	5.5±5

(B)	Female				Male
	OC user	Follicular phase	Luteal phase	Postmenopausal	
N	263	149	211	261	445
Age (years)	30.9±10	36.2±9	36.5±9	55±6	43.4±13
BMI (kg/m^2)	24.4±5	25.3±5	25.3±6	26.9±5	26.3±5
Waist circumference (cm)	81.8±12	84.7±13	84.4±14	90.7±13	96.2±14
North European ancestry % (Yes/No)	95/5	97/3	94/6	94/6	97/3
Area % (Amsterdam/Groningen/Leiden)	17/44/38	22/38/40	25/42/33	16/38/46	16/42/42
Frame % (General population/Primary care)	25/39/36	18/48/34	15/44/41	12/72/16	20/42/38
Education (years)	11.9±3	12.3±3	12.3±3	11.5±3	11.8±3
Partner % (Yes/No)	69/31	66/34	67/33	70/30	72/28
Smoking % (Current smoker/Former smoker/Never smoker)	37/27/37	36/28/36	46/29/25	29/46/25	43/34/23
Alcohol (drinks per week)	4±6	4.8±8	5.6±9	6.3±8	11±14
Recreational drug use in last month % (Yes/No)	10/90	5/95	7/93	2/98	9/91
Chronic disease % (Yes/No)	35/65	33/67	33/67	60/40	47/53
Anti-inflammatory medication % (Yes/No)	6/94	7/93	3/97	8/92	6/94
Lipid modifying agents % (Yes/No)	1/99	2/98	1/99	12/88	11/89
Antihypertensive medication % (Yes/No)	3/97	6/94	7/93	27/73	18/82
Physical activity (MET)	3763.3±3264	3703.6±2772	3340.5±2920	3749.2±2908	3728.3±3241
Systolic blood pressure (mm Hg)	126.8±13	126.7±17	124.9±19	143.1±22	144.1±20
Diastolic blood pressure (mm Hg)	77.8±9	79.6±11	77.7±13	85.6±11	84.5±12
IDS score	23.6±12	24.4±13	24.7±13	24±13	24.5±13

3.2.5 Simulation of group imbalance of sex and female hormonal status

In order to assess the effect of ignoring sex and female hormonal status on false discovery in biomarker studies, simulations were performed. In the first, an imbalance in the number of males and females in two groups was simulated. The sex composition was varied from having an even split of males and females in each group to having each group entirely composed of samples of one sex. For each simulation, 50 samples were selected with replacement from male and female NESDA control subjects to represent each group. A *t*-test was used to test for differences between the two groups in each of the 171 analyte levels and the number of false discoveries (p-value < 0.05) was recorded. Ten thousand simulations were performed for each male/female group composition and the average percentage of false discoveries was calculated. In the second simulation, this procedure was repeated to investigate false discoveries in two groups of females with an imbalance in the composition of reproductive-age OC users.

3.2.6 Classification of sex and female hormonal status

Data from NESDA control subjects were used to build a random forest classifier to distinguish between males, postmenopausal females, OC users, and females with a menstrual cycle. In addition to the concentrations of the 171 serum analytes, covariates used in the linear models were included, apart from recruitment method. Additional covariates waist circumference (centimetres), systolic blood pressure and diastolic blood pressure (mm Hg) were also used. The out-of-bag (OOB) error in this training sample was assessed and the remaining 1,329 NESDA samples were used to test the performance of the classifier. These test samples were largely comprised of samples from individuals diagnosed with a current, lifetime, or second year follow-up diagnosis of depressive and/or anxiety disorders. Demographic and other information for these remaining NESDA samples can be found in **Table 3.1B**. This procedure was repeated to construct a random forest classifier for female hormonal status only using age and a subset of 79 assays measured in previous versions of the Human DiscoveryMAP® multiplex immunoassay platform (see the **Appendix** (**Table A.1**) to see analytes measured in NESDA and previous cohorts). For both classifiers, 5000 trees were constructed using the randomForest packages in R (*329*). Mean decrease in the Gini index, a measure of node purity, was used to assess variable importance. **Chapter 2** provides more details on random forests.

3.3 Results

In the NESDA controls a number of demographic, lifestyle, and health variables differed significantly between males, postmenopausal females, females using OCs, and females in

the follicular and luteal phases of the menstrual cycle. These included differences in age, BMI, waist circumference, alcohol consumption, presence of chronic disease, use of antihypertensive medication, blood pressure, IDS scores, collection area, and recruitment method (**Table 3.1A**). With the exception of IDS scores and collection area, these differences were also observed in the remaining NESDA data (**Table 3.1B**). These variables were associated with changes in analyte levels. For example, the levels of 47 of the 171 analytes differed by recruitment method (whether participants were recruited from the community, general practice, or specialised mental health care facility*;* $p<0.05$). Age, BMI, collection area, recruitment method, use of antihypertensive medication, and weekly alcohol consumption were the variables most often selected in the stepwise regression of NESDA control data as additional covariates.

ComBat was effective in normalizing analyte concentrations across plates, as illustrated for selected molecules in **Figure 3.1**. The concentrations of 96 serum analytes differed between males and females ($q<0.05$) in the linear model described in **Section 3.2.4**. Of these, 43 had higher concentrations in females and 53 had higher concentrations in males (**Table 3.2**; **Figure 3.2**). Findings from the meta-analysis of *Ramsey et al (2012)* (*341*) were largely consistent with the current study, with 49 of the findings replicated out of 60 assays overlapping between the studies (**Table 3.2**; **Figure 3.2**).

Sex differences in 45 of the 96 serum analytes varied significantly ($q<0.05$) with female hormonal status (postmenopausal, OC use, follicular, and luteal phase of the menstrual cycle). An additional 21 serum analytes also showed significant ($q<0.05$) variation in their concentrations with female hormonal status (**Table 3.3**; **Figure 3.3**). Linear contrasts in the model revealed significant differences between females using OCs and those with menstrual cycles in the levels of 55 of these 66 serum analytes. Twenty-six of these analytes were different between postmenopausal females and females with a menstrual cycle. **Table 3.3** and **Figure 3.3** show that most of the significant differences between postmenopausal females and females with menstrual cycles opposed the differences in analyte levels between females with menstrual cycles and males. Only five of the 66 serum analytes differed significantly between females in the luteal and follicular phases of the menstrual cycle (**Table 3.3**). Follicle-stimulating hormone (FSH), neuropilin-1, and insulin-like growth factor binding proteins (IGFBP)-2 and -3 were increased in the follicular phase, while progesterone was increased in the luteal phase. The PCA plot in **Figure 3.4** was constructed using analyte data and further illustrates the differences in serum molecular profiles between males and females for each hormonal status. In the first principal component, OC users and males showed the greatest separation between groups, while females in the follicular and luteal phases of the menstrual cycle showed the smallest separation. Analytes that varied

significantly with sex and female hormonal status were related to metabolic processes, developmental processes, cell communication and signal transduction, transport, growth and cell proliferation, chemotaxis, response to stress, cell death, defence, inflammatory, and immune response, nervous system development, vasculature development, and others (**Table 3.2** and **Table 3.3**). In total, 45 of the 117 analytes that differed between males and females or with hormonal status were also reported to be altered in schizophrenia patients and 23 were changed in MDD patients.

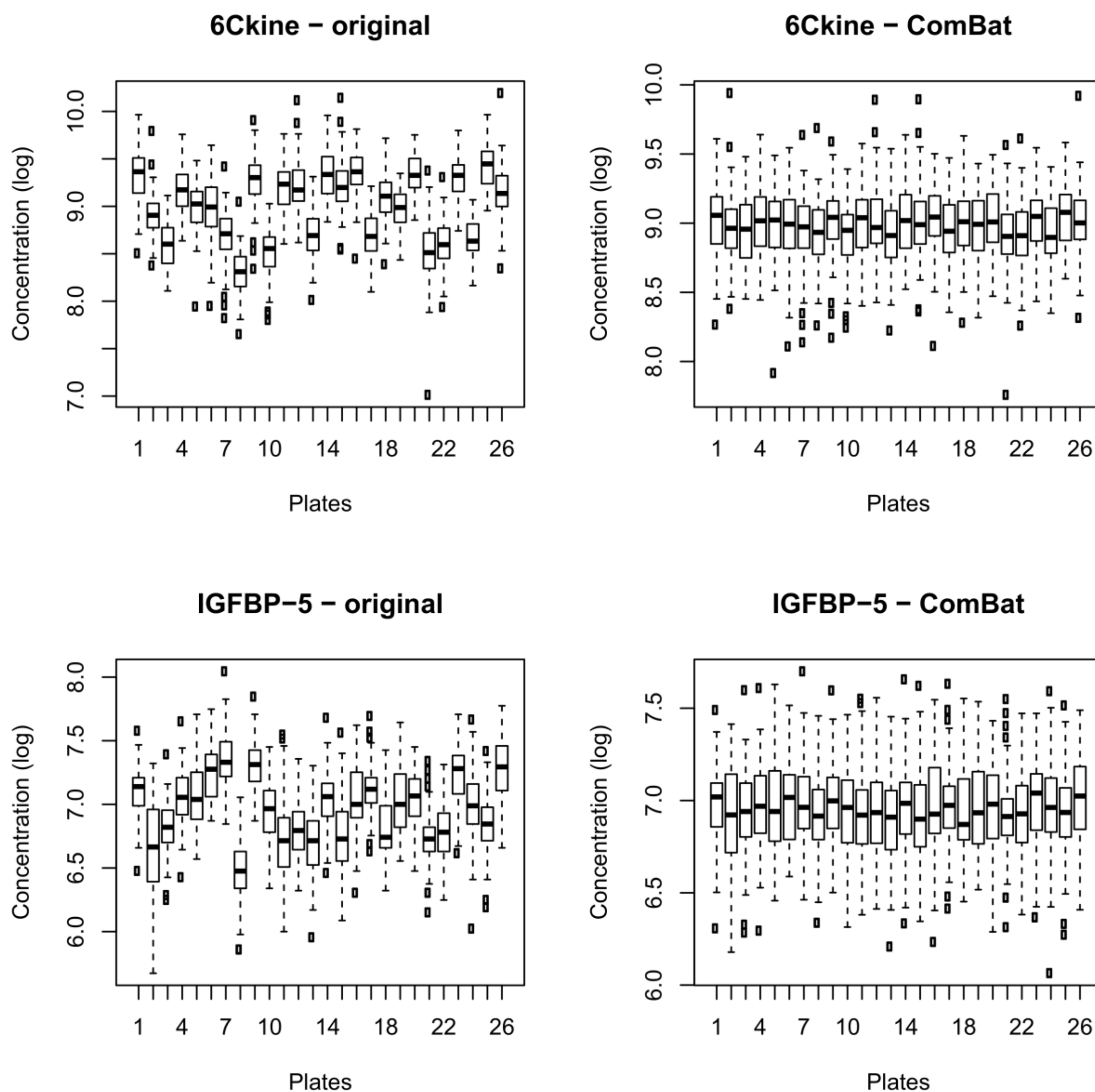


Figure 3.1. ComBat-normalized 6Ckine and insulin-like growth factor binding protein (IGFBP)-5 analyte concentrations. Boxplots across all plates are shown for log_2-transformed analyte concentrations before (left) and after (right) applying ComBat.

Table 3.2. Serum analytes elevated in (A) females and (B) males. Analytes are ordered from top to bottom by decreasing significance in the discovery cohort (reported in the columns of *q*-values). Robust regression was used where * follows the analyte name. The log_2 ratios of serum molecular concentrations in males compared to females shown in the plots were coefficients from the linear regression. Serum analytes that varied significantly with female hormonal status are indicated (see **Table 3.3**). These are coloured in blue where analytes were significantly different between oral contraceptive pill users and females with a menstrual cycle; orange where they were significantly different between postmenopausal females and females with a menstrual cycle; and yellow where there were multiple significant contrasts. Agreement with our previous study is shown by a ✓. SCZ and MDD markers are indicated by ↑ (elevated in patients); ↓ (reduced in patients); or ✗ (conflicting evidence for elevated and reduced levels in patients). Biological processes were found from gene ontology (GO) terms. **Abbreviations:** NM (not measured in our previous study); SCZ (schizophrenia); MDD (major depressive disorder); MP (metabolic process); DP (developmental process); CC & ST (cell communication and signal transduction); IR (defense/immune/inflammatory response); CP (cell proliferation); NSD (nervous system development); VD (vasculature development). Analyte abbreviations can be found in the **Appendix** (**Table A.1**). **GO terms (left to right):** GO:0008152; GO:0032502; GO:0007154 and GO:0007165; GO:0006810; GO:0040007 and GO:0008283; GO:0006935; GO:0006950; GO:0008219; GO:0006952 and GO:0006955 and GO:0006954; GO:0007399; GO:0001944.

(A)							Biological processes										
Analytes	log_2 ratio (male/female)	Q-value	Varies with hormonal status	Agrees with *Ramsey et al, 2012*	SCZ marker	MDD marker	MP	DP	CC & ST	Transport	Growth/CP	Chemotaxis	Response to stress	Cell death	IR	NSD	VD
Leptin	-2.06	4.E-64		✓	↑	↑	✓	✓	✓	✓	✓		✓	✓		✓	
FABP, adipocyte*	-0.68	1.E-31	✓	NM			✓	✓					✓		✓		
GH	-2.38	1.E-29	✓	✓	↓	↓	✓	✓	✓	✓	✓						
SHBG	-1.09	4.E-29	✓	✓				✓			✓						
Apo A-I*	-0.25	3.E-17		✓	↓	↓	✓	✓	✓	✓	✓	✓	✓		✓	✓	
Adiponectin	-0.60	8.E-15		✓	↓	↓	✓	✓	✓	✓	✓		✓	✓	✓		
TBG	-0.29	6.E-11	✓	✓	↑		✓	✓		✓							
IGFBP-1	-1.04	1.E-10	✓	NM			✓	✓	✓		✓		✓				
FSH*	-0.99	1.E-10	✓		↑		✓	✓	✓		✓						
CRP	-1.16	4.E-10	✓	✓	↑	↑	✓	✓		✓			✓		✓	✓	
uPAR*	-0.37	1.E-09		NM			✓		✓			✓	✓				
ErbB3*	-0.52	3.E-09	✓	NM			✓	✓	✓	✓	✓		✓	✓	✓	✓	
NT proBNP	-0.91	5.E-09	✓	✓			✓	✓	✓	✓	✓						✓
OPG*	-0.23	1.E-08	✓	NM				✓	✓					✓			
AAT*	-0.20	1.E-07	✓	✓	↑		✓			✓			✓		✓		
A2Macro	-0.21	2.E-07		✓		↓	✓	✓	✓	✓			✓		✓		
PRL*	-0.36	7.E-07						✓	✓	✓	✓						
ENA- 78*	-0.41	1.E-06		✓	↑	↓			✓		✓	✓	✓		✓		
BAFF*	-0.15	3.E-06		NM					✓	✓	✓				✓		
Factor VII*	-0.17	2.E-05	✓	✓	✗		✓	✓	✓			✓	✓				
CLU*	-0.09	6.E-05	✓				✓	✓	✓	✓	✓		✓	✓	✓	✓	
TFF3*	-0.17	1.E-04	✓	✓									✓		✓		
Transferrin	-0.11	3.E-04	✓	✓	↓	↓				✓			✓				
IgM	-0.27	7.E-04	✓	✓		↓											
VEGFR-3*	-0.27	8.E-04	✓	NM			✓	✓	✓		✓		✓	✓			✓
GRO-A*	-0.24	0.001		✓			✓	✓	✓		✓	✓	✓		✓	✓	
Fetuin A	-0.12	0.002	✓	✓	↓		✓	✓	✓	✓			✓		✓		
Collagen IV	-0.23	0.002	✓	NM			✓	✓	✓	✓	✓	✓		✓		✓	✓
MMP-7	-0.12	0.003		✓			✓			✓	✓		✓		✓		
Galectin-3*	-0.11	0.003		NM			✓	✓	✓	✓	✓	✓	✓	✓	✓		
MDC	-0.38	0.004			↑				✓			✓	✓		✓		
Haptoglobin	-0.14	0.004		✓	↑		✓			✓			✓	✓	✓		
THP*	-0.17	0.005		✓				✓			✓		✓		✓		
TSH*	-0.23	0.008				↓	✓		✓		✓						
TG*	-0.27	0.013		NM			✓	✓	✓	✓						✓	
uPA	-0.13	0.016	✓	NM			✓		✓		✓	✓	✓				
Apo A-II*	-0.08	0.016	✓	NM	↓		✓	✓	✓	✓			✓		✓		
LOX-1*	-0.20	0.018	✓	✓			✓						✓	✓	✓		
M-CSF*	-0.15	0.018	✓				✓	✓	✓		✓		✓		✓		
Vitronectin	-0.08	0.029	✓	✓			✓	✓	✓	✓			✓		✓		
Resistin	-0.15	0.032			↓		✓	✓	✓	✓	✓						
IL-12p40*	-0.10	0.049		✓	↓		✓	✓	✓	✓	✓		✓	✓	✓		
C-Peptide*	-0.11	0.050	✓				✓	✓	✓	✓	✓		✓	✓	✓		

(B)							Biological processes										
Analytes	$\log_2$ ratio (male/female)	Q-value	Varies with female hormonal status	Agrees with *Ramsey et al, 2012*	SCZ marker	MDD marker	MP	DP	CC & ST	Transport	Growth/CP	Chemotaxis	Response to stress	Cell death	IR	NSD	VD
Testosterone*	1.98	3.E-174	✓	✓													
MMP- 3	0.99	2.E-46		✓			✓		✓				✓	✓			
FRTN	1.65	4.E-27	✓	✓	↑	↑				✓	✓				✓		
HER- 2*	0.38	1.E-14	✓	NM			✓	✓	✓		✓	✓	✓		✓	✓	✓
Eotaxin-1	0.58	8.E-13		✓			✓	✓	✓		✓	✓	✓		✓		✓
AXL*	0.33	5.E-10	✓	✓			✓	✓	✓	✓			✓	✓	✓	✓	
VKDPS*	0.19	3.E-09		✓			✓			✓			✓		✓		
tPA	0.33	4.E-09	✓	NM			✓		✓	✓			✓				
NrCAM	0.40	4.E-09		NM	↓			✓	✓		✓	✓				✓	✓
Myoglobin*	0.34	1.E-08		✓				✓					✓				
TM*	0.20	1.E-08	✓	✓									✓				
TTR	0.17	1.E-07	✓	✓	↓		✓		✓	✓							
PSAT*	0.43	2.E-07		NM			✓										
E-Selectin	0.35	4.E-07	✓	✓			✓		✓	✓			✓		✓		
IL-6rB*	0.11	9.E-07	✓	NM			✓	✓	✓		✓		✓	✓	✓	✓	
Osteopontin	0.41	1.E-06	✓	✓				✓			✓	✓	✓		✓	✓	
CK-MB	0.45	2.E-06		✓			✓	✓								✓	
IGFBP-6*	0.17	4.E-06		NM			✓		✓		✓						
SAP*	0.21	5.E-06	✓	✓	↑	↑	✓	✓					✓		✓		
Apo D	0.18	6.E-06	✓	✓	✗		✓	✓	✓	✓	✓		✓		✓	✓	✓
HCC-4*	0.29	1.E-05	✓	✓					✓			✓	✓		✓		
Prostasin*	0.22	2.E-05		NM			✓			✓							
Angiogenin	0.15	7.E-05	✓	NM			✓	✓	✓	✓	✓		✓				✓
ALDR*	0.14	8.E-05		NM			✓						✓				
Cathepsin D	0.16	1.E-04		NM			✓										
HGF	0.26	1.E-04					✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
FAS*	0.22	2.E-04		✓			✓	✓	✓				✓	✓	✓		
Eotaxin-2	0.39	2.E-04		NM			✓	✓	✓		✓	✓	✓		✓		✓
Apo H*	0.14	2.E-04	✓	✓	↑		✓	✓		✓	✓		✓	✓			✓
PGI*	0.17	4.E-04		NM			✓										
MIP-1B	0.26	5.E-04	✓	✓	↑	↑	✓		✓	✓		✓	✓		✓		
A1Micro	0.10	5.E-04		✓			✓		✓	✓			✓		✓		
IL-18	0.26	5.E-04			↑		✓	✓	✓	✓	✓		✓		✓		✓
ACE	0.16	6.E-04	✓			↓	✓	✓		✓	✓						
CEA*	0.35	6.E-04		✓	↑			✓						✓			
MCP-1*	0.21	7.E-04		✓	↑		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
LGL*	0.27	8.E-04		NM			✓	✓						✓			
Gelsolin	0.12	0.001		NM				✓	✓	✓	✓		✓	✓		✓	
CA 15-3	0.26	0.002		NM			✓		✓				✓	✓			
GSTA	0.44	0.002			↑		✓	✓									
NSE*	0.21	0.003		NM			✓										
IL-6r	0.14	0.005			↑		✓	✓	✓		✓	✓	✓	✓	✓		
MCP-4	0.18	0.008	✓	✓				✓	✓			✓	✓		✓		
Cystatin-C	0.06	0.018		✓			✓	✓					✓		✓		
CgA*	0.17	0.021	✓		↑												
SCF*	0.09	0.022			✗		✓	✓	✓		✓		✓	✓	✓		
TSP1	0.12	0.023		✓	↓		✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
6Ckine	0.07	0.024		NM			✓	✓	✓	✓		✓	✓	✓	✓		
MCP-2*	0.13	0.032					✓		✓	✓		✓	✓		✓		
IGFBP-2	0.15	0.035	✓	✓	↑		✓	✓	✓		✓						
Tetranectin*	0.08	0.038		NM			✓	✓									
SOD-1*	0.15	0.039		✓		↑	✓	✓	✓	✓	✓		✓	✓		✓	
MIF*	0.18	0.043			↑	↑	✓	✓	✓	✓	✓	✓	✓	✓	✓		

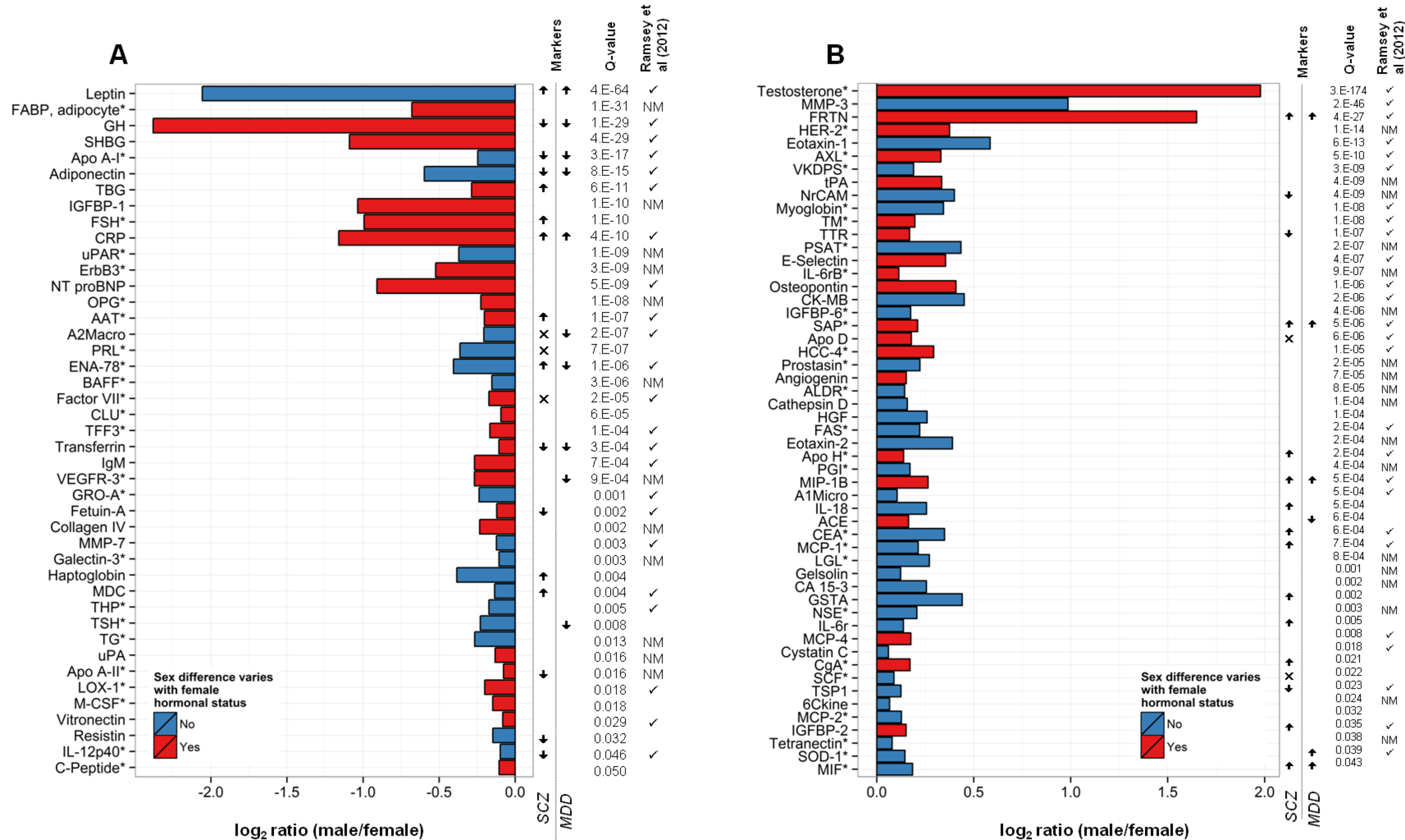


Figure 3.2. Serum analytes elevated significantly in (A) females and (B) males. Analytes are ordered from top to bottom by decreasing significance in the discovery cohort (reported in the columns of *q*-values). Robust regression was used where * follows the analyte name. The $\log_2$ ratios of serum molecular concentrations in males compared to females shown in the plots were coefficients from the linear regression. Analytes in the plots are coloured in red where sex differences varied significantly with female hormonal status (whether females were postmenopausal, in the luteal or follicular phase of the menstrual cycle, or taking the oral contraceptive pill). NM was used to indicate that the analyte was not measured in our previous study. To the right of the plots, biomarkers of schizophrenia (SCZ) and major depressive disorder (MDD) indicated by 🡡 (elevated in patients); 🡣 (reduced in patients); or ✕ (conflicting evidence for elevated and reduced levels in patients). Analyte abbreviations can be found in the **Appendix** (**Table A.1**).

Table 3.3. Serum analytes varying significantly with female hormonal status. Analytes are ordered from top to bottom by decreasing significance in the discovery cohort (reported in the column of ANOVA *q*-values). Robust regression was used where * follows the analyte name. The log_2 ratios of serum molecular concentrations were coefficients from the linear regression. SCZ and MDD markers are indicated by ↑ (elevated in patients); ↓ (reduced in patients); or × (conflicting evidence for elevated and reduced levels in patients). Biological processes were found from gene ontology (GO) terms. Significant differences in analyte concentrations between groups are highlighted. **Abbreviations:** Q (*q*-value); R (log_2 ratio); SCZ (schizophrenia); MDD (major depressive disorder); MP (metabolic process); DP (developmental process); CC & ST (cell communication and signal transduction); IR (defense/immune/inflammatory response); CP (cell proliferation); NSD (nervous system development); VD (vasculature development). Analyte abbreviations can be found in the **Appendix (Table A.1). GO terms (left to right):** GO:0008152; GO:0032502; GO:0007154 and GO:0007165; GO:0006810; GO:0040007 and GO:0008283; GO:0006935; GO:0006950; GO:0008219; GO:0006952 and GO:0006955 and GO:0006954; GO:0007399; GO:0001944.

Analytes	ANOVA	Follicular phase / luteal phase		Oral contraception / menstrual cycle		Postmenopausal / menstrual cycle		Males / females with menstrual cycle		SCZ marker	MDD marker	Biological processes										
	Q	R	Q	R	Q	R	Q	R	Q			MP	DP	CC & ST	Transport	Growth / CP	Chemotaxis	Response to stress	Cell death	IR	NSD	VD
LH*	3.E-135	0.27	0.17	-0.46	3.E-06	2.03	6.E-39	-0.24	0.024	↑		✓	✓	✓								
FSH*	2.E-78	0.69	7.E-04	-0.93	4.E-09	2.68	2.E-39	-0.50	4.E-05	↑		✓	✓	✓		✓						
TFF3*	2.E-65	0.10	0.32	3.43	2.E-48	0.01	0.97	-0.15	0.008									✓		✓		
TBG	8.E-24	0.01	0.98	0.48	1.E-18	-0.18	0.008	-0.16	0.001	↑		✓	✓		✓							
AAT*	8.E-20	0.14	0.055	0.43	1.E-16	-0.03	0.78	-0.05	0.31	↑		✓			✓			✓		✓		
SHBG	1.E-16	0.21	0.32	0.86	8.E-12	-0.21	0.16	-0.83	1.E-13				✓			✓						
ErbB3*	2.E-15	-0.32	0.12	0.91	1.E-11	0.11	0.54	-0.18	0.15			✓	✓	✓	✓	✓		✓	✓	✓	✓	
Osteopontin	2.E-15	0.14	0.46	-0.61	1.E-07	0.64	7.E-06	0.38	2.E-04				✓			✓	✓	✓		✓	✓	
SAP*	1.E-14	-0.03	0.85	0.49	2.E-14	0.18	0.025	0.44	1.E-16	↑	↑	✓	✓					✓		✓		
Testosterone*	4.E-13	-0.11	0.24	-0.32	2.E-08	0.18	0.012	1.91	2.E-137													
Progesterone	6.E-12	-0.72	5.E-08	-0.44	2.E-06	-0.27	0.020	-0.22	0.010	↑												
Transferrin	3.E-11	-0.08	0.25	0.27	5.E-11	-0.02	0.81	-0.01	0.85	↓	↓				✓			✓				
IGFBP-2	1.E-10	0.33	0.025	-0.65	8.E-11	-0.11	0.43	-0.12	0.24	↑		✓	✓	✓		✓						
HER-2*	9.E-10	-0.02	0.95	-0.34	2.E-06	0.28	0.002	0.34	4.E-08			✓	✓	✓		✓	✓	✓		✓	✓	✓
CRP	4.E-09	0.22	0.66	1.45	4.E-08	-0.06	0.91	-0.64	0.008	↑	↑	✓	✓		✓			✓		✓	✓	
Vitronectin	4.E-09	0.06	0.51	0.29	3.E-08	-0.07	0.33	0.00	0.98			✓	✓	✓	✓			✓		✓		
Factor VII*	5.E-07	0.02	0.89	0.33	4.E-08	0.16	0.017	0.00	0.99	×		✓	✓	✓			✓	✓				
TTR	7.E-07	-0.13	0.056	0.23	1.E-07	0.13	0.006	0.30	2.E-13	↓		✓		✓	✓							
MCP-4	4.E-06	0.03	0.89	-0.39	7.E-05	0.22	0.072	0.10	0.30				✓	✓			✓	✓		✓		
OPG*	5.E-06	0.14	0.11	0.34	1.E-06	0.09	0.19	-0.09	0.11				✓	✓					✓			
CLU*	6.E-06	0.02	0.78	0.18	1.E-06	0.11	0.014	0.00	0.99			✓	✓	✓	✓	✓		✓	✓	✓	✓	
C3	7.E-06	-0.01	0.98	0.24	3.E-07	0.09	0.081	0.10	0.017	↑	↑	✓	✓	✓	✓			✓		✓		✓
Fetuin-A	8.E-06	-0.06	0.51	0.27	2.E-06	0.00	0.99	-0.02	0.81	↓		✓	✓	✓	✓			✓		✓		
M-CSF*	1.E-05	-0.13	0.32	0.43	3.E-06	0.04	0.78	0.02	0.85			✓	✓	✓		✓		✓		✓		
uPA	9.E-05	0.11	0.41	-0.34	2.E-05	-0.07	0.47	-0.28	7.E-05			✓		✓		✓	✓	✓				
VEGFR-2*	2.E-04	-0.04	0.67	0.22	1.E-05	0.07	0.30	0.11	0.009			✓	✓	✓		✓	✓		✓			✓
CgA*	2.E-04	-0.09	0.64	-0.42	2.E-04	0.16	0.32	0.07	**0.51**	↑												

AXL*	2.E-04	0.15	0.30	-0.02	0.87	0.29	0.001	0.41	8.E-08			✓	✓	✓	✓			✓	✓	✓	✓	
IGFBP-3*	2.E-04	0.08	0.034	-0.07	0.026	0.06	0.11	-0.03	0.31			✓	✓	✓		✓			✓			
VEGFR-3*	2.E-04	-0.12	0.52	0.47	5.E-05	0.01	0.98	-0.10	0.37			✓	✓	✓		✓		✓	✓			✓
Apo A-II*	3.E-04	-0.02	0.91	0.21	4.E-05	0.14	0.008	0.04	0.41	↓		✓	✓	✓	✓			✓		✓		
MSP*	3.E-04	0.00	1.00	0.39	1.E-04	0.28	0.004	0.18	0.017				✓									
IgM	5.E-04	0.27	0.15	0.16	0.24	-0.43	0.003	-0.34	0.002		↓											
Lp(a)	5.E-04	0.80	0.071	-0.87	0.004	0.23	0.54	-0.48	0.082		↑	✓			✓							
IGFBP-1	5.E-04	0.35	0.36	0.95	8.E-05	0.29	0.33	-0.59	0.006			✓	✓	✓		✓		✓				
VDBP	5.E-04	-0.23	0.15	0.40	2.E-04	0.04	0.85	0.07	**0.51**			✓			✓							
MMP-9*	8.E-04	0.05	0.83	0.25	0.008	-0.12	0.28	0.03	0.85	↑	↑	✓	✓	✓	✓		✓		✓		✓	
VCAM-1	9.E-04	0.08	0.37	-0.17	0.002	0.02	0.80	0.00	0.99	↑	↑	✓	✓	✓		✓	✓	✓		✓		
tPA	0.001	0.00	0.99	-0.28	8.E-04	0.08	0.52	0.26	7.E-04			✓		✓	✓			✓				
Fib 1C*	0.002	0.08	0.41	-0.15	0.007	0.04	0.55	0.00	0.99				✓									
Neuropilin-1	0.002	0.18	0.021	-0.14	0.012	0.01	0.91	0.02	0.83			✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Apo D	0.003	-0.11	0.24	-0.20	0.002	-0.16	0.026	0.06	0.32	✕		✓	✓	✓	✓	✓		✓		✓	✓	✓
TM*	0.003	0.08	0.32	-0.09	0.091	0.15	0.025	0.21	4.E-06									✓				
Insulin	0.005	0.01	0.99	0.62	4.E-04	0.23	0.27	0.29	0.082	↑	↑	✓	✓	✓	✓	✓		✓	✓	✓		
Angiogenin	0.005	0.01	0.98	0.20	6.E-04	0.14	0.063	0.27	1.E-07			✓	✓	✓	✓	✓		✓				✓
IGFBP-4*	0.005	0.00	0.99	0.17	5.E-04	0.10	0.11	0.03	0.61			✓	✓	✓		✓		✓		✓		
HCC-4*	0.008	0.06	0.81	0.34	0.003	0.32	0.008	0.51	3.E-08					✓			✓	✓		✓		
ANG-2*	0.009	0.00	1.00	-0.39	0.002	-0.08	0.59	-0.17	0.11	↑			✓	✓			✓	✓				✓
NT proBNP	0.010	0.30	0.44	0.00	0.99	-0.87	0.002	-1.16	2.E-08			✓	✓	✓	✓	✓						✓
Apo C-III	0.010	-0.15	0.25	0.24	0.003	0.19	0.055	0.07	0.40	↑	↑	✓		✓	✓			✓		✓		
IL-1ra	0.011	-0.06	0.66	0.23	0.002	-0.01	0.96	0.06	**0.41**	↑	↑	✓	✓	✓	✓			✓	✓	✓		
FRTN	0.013	-0.02	0.99	0.06	0.89	0.86	0.002	1.92	1.E-21	↑	↑				✓	✓				✓		
Angiotensinogen	0.013	0.22	0.84	1.44	0.002	0.24	0.74	0.25	0.64			✓	✓	✓	✓	✓		✓	✓	✓	✓	✓
E-Selectin	0.015	-0.17	0.32	-0.28	0.009	0.00	0.99	0.25	0.009			✓		✓	✓			✓		✓		
Apo H*	0.016	0.00	0.99	0.07	0.30	0.23	0.002	0.24	1.E-05	↑		✓	✓		✓	✓		✓	✓			✓
MIG*	0.016	0.22	0.28	0.33	0.012	0.27	0.11	0.15	0.22			✓	✓	✓	✓	✓	✓	✓		✓		
HGF receptor*	0.020	0.01	0.96	0.17	0.006	0.14	0.025	0.09	0.056			✓	✓	✓	✓	✓	✓	✓	✓		✓	
IL-6rB*	0.021	0.00	0.99	-0.02	0.62	0.12	0.010	0.14	2.E-05			✓	✓	✓		✓		✓	✓	✓	✓	
C-Peptide*	0.023	0.01	0.98	0.23	0.003	0.14	0.15	0.02	0.89			✓	✓	✓	✓	✓		✓	✓	✓		
LOX-1*	0.024	-0.01	0.99	0.17	0.25	-0.25	0.064	-0.21	0.055			✓						✓	✓	✓		
ACE	0.031	-0.10	0.41	-0.10	0.20	0.13	0.14	0.16	0.017		↓	✓	✓		✓	✓						
GH	0.031	-0.13	0.86	0.68	0.032	-0.24	0.55	-2.19	3.E-16	↓	↓	✓	✓	✓	✓	✓						
SDF-1*	0.032	-0.04	0.52	-0.10	0.016	-0.10	0.025	-0.04	0.25				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MIP-1B	0.034	0.27	0.13	-0.26	0.026	0.04	0.85	0.18	0.10	↑	↑	✓		✓	✓		✓	✓		✓		
FABP, adipocyte*	0.037	0.16	0.19	0.12	0.22	0.24	0.020	-0.57	4.E-16			✓	✓					✓		✓		
Collagen IV	0.039	0.32	0.064	-0.15	0.23	-0.27	0.034	-0.37	3.E-04			✓	✓	✓	✓	✓	✓		✓		✓	✓

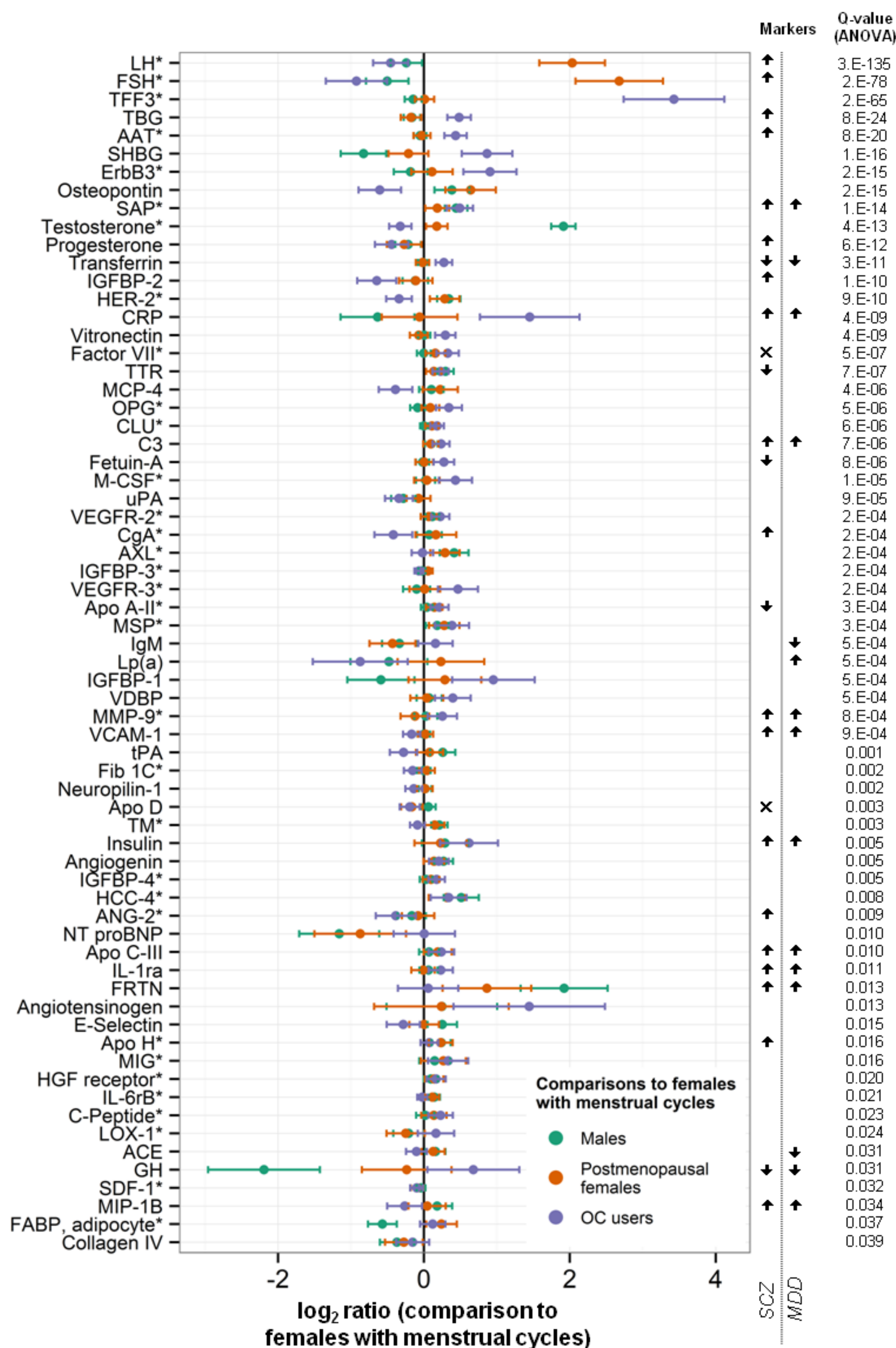
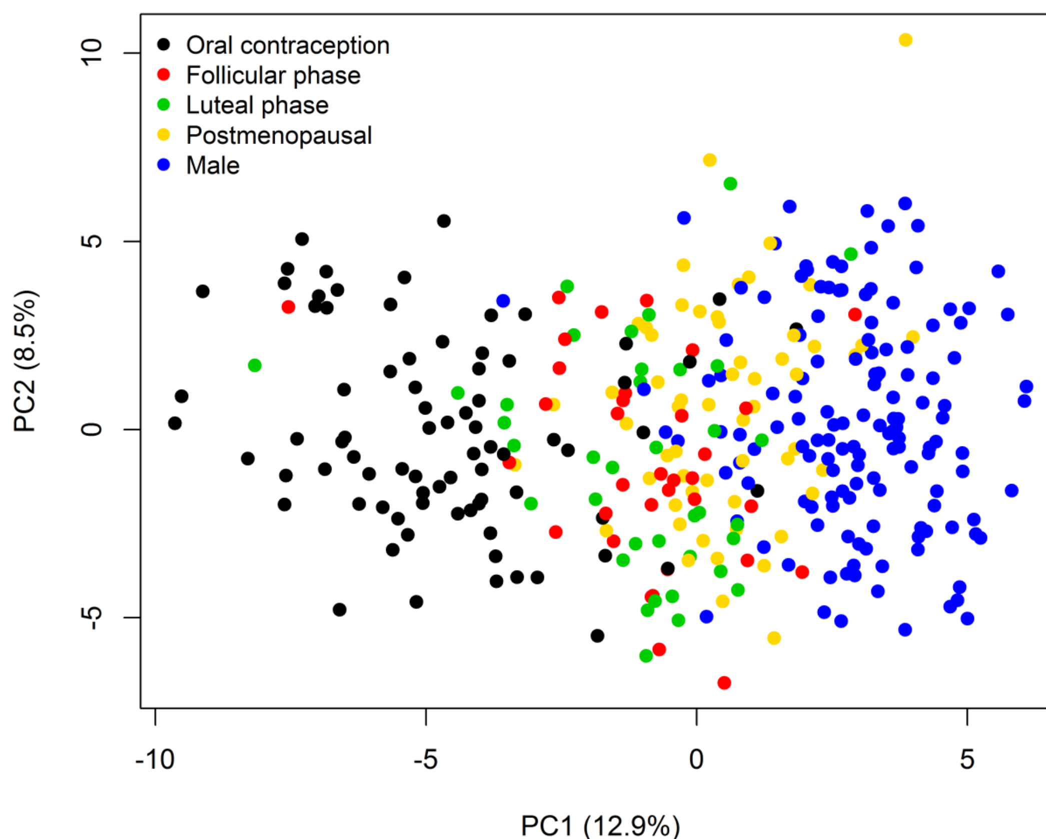


Figure 3.3. Serum analytes varying significantly with female hormonal status. Analytes are ordered from top to bottom by decreasing significance in the discovery cohort (reported in the column of *q*-values). Robust regression was used where * follows the analyte name. The $\log_2$ ratios of serum molecular concentrations of males, postmenopausal females, and females taking the oral contraceptive pill compared to females with a menstrual cycle, shown in the plot with their 95% confidence intervals (FDR-adjusted), were coefficients from the linear regression. To the right of the plot, biomarkers of schizophrenia (SCZ) and major depressive disorder (MDD) are indicated by 🡑 (elevated in patients); 🡓 (reduced in patients); or ✕ (conflicting evidence for elevated and reduced levels in patients). Analyte abbreviations can be found in the **Appendix** (**Table A.1**).

Figure 3.4. **Principal component analysis (PCA) plot of NESDA control samples.** The first two principal components (PCs) are plotted and coloured according to sex and female hormonal status. PCA was performed using all analyte data and percentage of variation accounted for by each principal component is included in brackets with the axis labels.



Simulations showed that analyte data with an imbalanced representation of males and females yielded a false discovery rate of up to 39.3%, as seen in **Table 3.4**. Furthermore, when the proportion of OC users differed between groups of pre-menopausal females, up to 41.5% of measured analytes were significantly different. For comparison, balanced groups had a false discovery of 4.4% of analytes on average. By constructing a random forest classifier from covariate and analyte data, correct classification of samples as male, postmenopausal female, OC user, or female with a menstrual cycle could be made for 94.5% of the control data (OOB samples) and 92.0% of test samples from the remaining NESDA cohort (depressive/anxiety disorders) (**Table 3.5A**). A similar performance was achieved (86.5% out-of-bag and 87.6% test accuracy) classifying females with menstrual cycles, OC users, and postmenopausal females using a subset of the variables and analyte measurements collected in previous experiments (**Table 3.5B**). Testosterone, FSH, trefoil factor 3 (TFF3), luteinizing hormone (LH), matrix metalloproteinase-3 (MMP-3), sex hormone-binding globulin (SHBG), leptin, age, ferritin, and thyroxine binding globulin (TBG)

were the ten most important variables for classification of sex and female hormonal status using the full set of analytes and demographic and other variables (**Figure 3.5**).

Table 3.4. **Average percentage of false discoveries in groups with imbalances in sex and female hormonal status.** Each group contained 50 individuals sampled with replacement. In the first set of trials, the effect of imbalances between the number of males and females in each group on the false discovery of differences in analyte means between groups ($p<0.05$) were tested. The first column shows the percentage of males in each group and the second column shows the mean percentage of analytes falsely discovered from a total of 10,000 simulations. In the second set of trials, the effect of imbalances in groups composed of reproductive-age women between the number of females using oral contraceptives (OC) and females with a menstrual cycle on the false discovery of differences in analyte means between groups were tested. The first column shows the percentage of OC users in each group and the third column shows the mean percentage of analytes falsely discovered from a total of 10,000 simulations. Significant differences in analyte means between groups were found using Welch's *t*-test.

% Group 1:Group 2 Male/OC users	False discoveries % Males vs Females	False discoveries % OC users vs Menstrual cycle
50:50	4.4	4.4
60:50	4.9	5.0
40:50	4.9	4.9
60:40	6.6	6.7
70:50	6.7	6.8
30:50	6.7	6.7
20:50	9.8	10.2
80:50	10.1	10.2
70:30	13.6	14.4
10:50	13.6	14.7
90:50	14.3	14.7
0:50	18.0	19.7
100:50	19.0	19.8
80:20	22.6	24.2
90:10	31.4	33.6
100:0	39.3	41.5

Table 3.5A-B. Test error for random forest classifiers. (A) shows the error in the test set for the classifier constructed to predict sex and female hormonal status using the full set of covariates and analyte measurements. **(B)** shows the error in the test set for the classifier constructed to predict oral contraceptive (OC) users, postmenopausal females, and females with a menstrual cycle using age and a reduced set of analyte measurements. The out-of-bag (OOB) error is shown for the training data (control subjects).

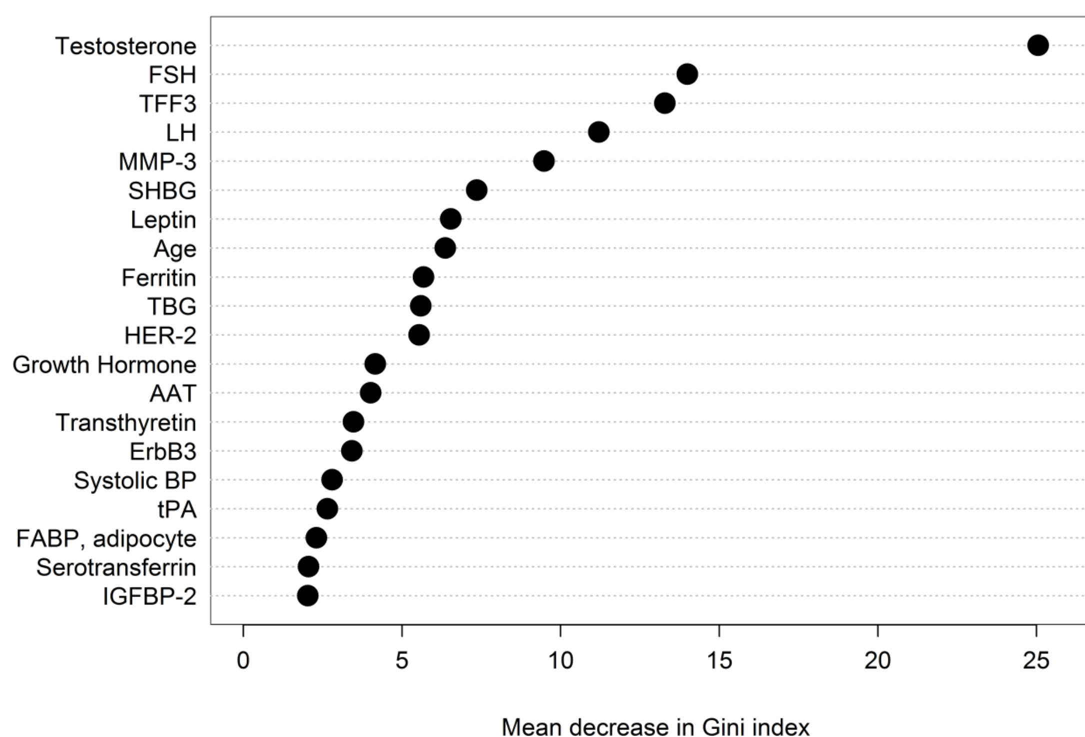
A

Actual class	Predicted class				**Error (%)**	**OOB error (%)**
	OC	Menstrual cycle	Postmenopausal	Male		
OC	220	32	7	4	16.3	7.6
Menstrual cycle	20	318	14	8	11.7	12.7
Postmenopausal	1	16	241	3	7.7	5.3
Male	0	1	0	444	0.2	0.7

B

Actual class	Predicted class			**Error (%)**	**OOB error (%)**
	OC	Menstrual cycle	Postmenopausal		
OC	220	36	7	16.3	13.9
Menstrual cycle	37	312	11	13.3	19.7
Postmenopausal	2	17	242	7.2	5.3

Figure 3.5. **Plot of random forest variable importance for classifying sex and female hormonal status**. This figure shows variable importance, measured as the mean decrease in Gini index, for the top 20 variables for the random forest classifier constructed to predict sex and female hormonal status using the full set of covariates and serum analyte measurements. **Abbreviations:** BP (blood pressure). Analyte abbreviations can be found in the **Appendix** (**Table A.1**).



3.4 Discussion

This chapter presents a comprehensive investigation of variations with sex and female hormonal status in the serum concentrations of a large panel of proteins and small molecules in individuals not affected by mental disorders. A previous meta-analysis by *Ramsey et al (2012)* (*341*) of serum molecular sex differences did not evaluate the effect of female hormonal status or take into account many potentially gendered demographic, lifestyle, and health variables. The extensive characterization of the NESDA cohort allowed analyses to account for differences in age, BMI, alcohol consumption, IDS scores, and other covariates among males, postmenopausal females, OC users, and females in the luteal and follicular phases of their menstrual cycles.

In total, the serum concentrations of 117 molecules were associated with sex and/or female hormonal status and these had diverse functions in immune, cardiovascular, metabolic, nervous, and endocrine pathways. Almost half of the 96 molecules found to have different concentrations between males and females also varied significantly with female hormonal status. Future studies should therefore use caution when interpreting sex differences in the concentrations of these and in the 21 additional molecules that varied with female hormonal status. In particular, extensive differences in serum molecular concentrations of OC users were identified compared to females with menstrual cycles. In the *Ramsey et al (2012)* (*341*) meta-analysis for example, the magnitude and/or significance of sex differences reported in the serum concentrations of TFF3, α1-antitrypsin (AAT), CRP, fetuin A, serotransferrin, TBG, SHBG, and a number of others were likely due to OC use by females in the study. The extensive differences in serum molecular profiles between reproductive-age females with menstrual cycles and using OCs was reflected in the high rates of false discovery of significantly different analyte concentrations between groups with imbalanced compositions of OC users and females with menstrual cycles. These findings may provide insight into processes involved in the reduced all-cause mortality (*347*), lower risk of endometrial and ovarian cancer and bacterial pelvic inflammation (*348*), higher risk of certain cardiovascular events (*348*), higher risk of Crohn's disease and ulcerative colitis (*349*), and potential effects on symptoms of depression (*350*) in OC users. Deviations in molecular serum concentrations in postmenopausal females were found that tended towards the levels found in males. These included TBG that binds thyroid hormone; osteopontin, involved in bone remodelling and immune functions; immunoglobulin M (IgM), an antibody made in response to first exposure to an antigen; and others. This likely obscured and/or reduced the magnitude of the sex differences found in the *Ramsey et al (2012)* (*341*) meta-analysis, further emphasizing the need to consider hormonal status in analyses of sex differences. Serum molecules that differ in postmenopausal females may play a role in the increased

prevalence of a number of conditions after menopause, including cardiovascular disease (*351*), osteoporosis (*352*), metabolic syndrome (*353*), Alzheimer's disease and late-onset schizophrenia (*354*), and others. On the other hand, few differences were observed between females in the follicular and luteal phases of their menstrual cycles. However, the considerable variability in hormonal levels within phases and particularly prior to ovulation, timing of phases between individuals, and self-reporting of menstrual cycle status likely contributed to the lack of findings.

Sex hormones and hormones involved in reproductive processes have been implicated in schizophrenia, MDD, and other mental disorders. These include estrogen, testosterone, progesterone, FSH, and LH (*209*, *267–269*). Imbalances in levels of these hormones may affect the function of endocrine, immune, metabolic, cardiovascular, and other pathways in a sex-dependent manner. Progesterone, testosterone, and estrogen are known to modulate immune response (*53*). The levels of proteins associated with acute inflammatory response like C-reactive protein (CRP), serum amyloid P-component (SAP), AAT, and interleukin-1 receptor antagonist (IL-1ra) were elevated in OC users compared to females with menstrual cycles. However, other inflammatory proteins, such as E-selectin, eotaxins -1 and -2, monocyte chemotactic protein (MCP) -1, -2, and -4, and interleukin (IL) -6 receptor and -6 receptor subunit β were reduced in females compared to males. Ferritin and osteopontin levels were lower in females with menstrual cycles and OC users compared to males and increased after menopause, further illustrating the relationship between inflammatory markers and reproductive hormones. Many of these serum molecules, including CRP, SAP, AAT, ferritin, IL-1ra, IL-6 receptor, and MCP-1, were also reported to be elevated in schizophrenia and/or MDD patients. Inflammatory processes are hypothesized to play an important role in these and other mental disorders (*31*, *355*, *356*) and these results raise the possibility of a sex-dependent component to these. Levels of serum molecules involved in metabolic processes also varied with sex and hormonal status and are reported to be disturbed in mental disorders (*49*, *51*, *132*, *207*, *357*, *358*). This included serum levels of apolipoproteins AI, AII, CIII, D, and H involved in lipid transport, which were also reported to be altered in patients with MDD and/or schizophrenia diagnoses. Furthermore, sex and female hormonal variations in insulin-like growth factor (IGF)-1 activity may have implications for sex-dependent metabolic alterations in mental disorders. Insulin resistance has been reported in MDD and schizophrenia (*132*, *359*). IGF-1 deficiency may play a substantial role in early stage schizophrenia (*359*) and IGF-1 has been observed to have antidepressant-like effects in rats (*360*). Although IGF-1 was not measured in this study, serum levels of insulin, C-peptide, growth hormone, and insulin-like growth factor-binding protein (IGFBP)-1, -2, -3, -4, and -6 differed with sex and hormonal status. Evidence was also found for sex and

hormonal status influencing serum concentrations of analytes involved in vasculature development, including angiogenin and neuropilin-1 and components of the renin-angiotensin system, angiotensinogen and angiotensin-converting enzyme (ACE). These may be relevant to MDD and schizophrenia, as abnormalities in the renin-angiotensin system (*361*), angiogenesis (*362*, *363*), and brain blood flow (*115*, *197*) have been observed in these conditions. Interactions between metabolic, inflammatory, and cardiovascular processes have the potential to further result in a complex host of sex-dependent pathophysiological mechanisms in mental disorders. However, the implications of these findings for sex differences in brain function remain to be seen.

Results from this investigation show that accounting for sex and female hormonal status may help to avoid confounding with a variable of interest. Although many biomarker and other studies currently match cases and controls for sex, some ignore sex, do not report it, and/or do not account for it. For example, a highly-cited study by *James et al (2004)* (*364*) of metabolic biomarkers found evidence of increased oxidative stress in autism, but reported only the proportion of boys in the autistic group (70%) and not in the control group. The simulations performed in this study support the practice of matching for sex, with the most severe discrepancy in the proportion of males and females between two groups yielding almost 40% false positive results. Less severe imbalances also yielded high numbers of false positives. For instance, a ratio of 70% males in one group to 30% males in the other generated 13.6% false positive results. These discrepancies are more likely in studies of diseases and mental disorders with a higher prevalence in one sex, such as MDD or autism spectrum disorders. Female hormonal status was also an important variable to consider and one for which most biomarker studies do not account, unless they are specifically designed to investigate sex effects. Simulations yielded up to 40% false positives where the composition of two groups made up of OC users and females with menstrual cycles were uneven. Group imbalances in OC use may be more likely when investigating mental disorders, as depressed female adults and adolescents have been reported to be more likely to have unprotected sex (*365*) and the consistency of contraceptive use was found to be reduced in female veterans with mental illness, particularly among those with comorbid substance use disorders (*366*). Apart from reducing the number of false discoveries in biomarker studies, accounting for sex and female hormonal status can also increase the power of analyses by accounting for variability in analyte concentrations. In light of these considerations, high accuracy classification tools were developed here to assist in labelling samples as males, postmenopausal females, OC users, and females with a menstrual cycle where this information has not been collected. This classification tool was effective in labelling subjects free of mental illness and also those with a current, lifetime, or second year

diagnosis of depressive and/or anxiety disorders. It was also effective in classifying hormonal status of female subjects when a limited number of serum analytes and other variables, measured in previous versions of the multiplex immunoassay platform, were used as predictors. This may provide a means of re-evaluating past cohorts accounting for female hormonal status.

In this study estrogen dose, progestogen type, brand, and monophasic or triphasic OC use were not recorded. However, Dutch medical guidelines recommend that if a combined OC is chosen, preference is for one with levonorgestrel and 30µg of ethynyl estradiol (*367*). OCs with levonorgestrel were used by the largest proportion of subjects (57%) in the control arm of a recent study on venous thrombosis in the Netherlands and progestogen-only pills were used by the smallest proportion of subjects (0.6%) (*368*). In addition, active-pill or hormone-free interval for OC users, duration of OC use, and timing of menopause were also not recorded and would need to be taken into account in future studies. Other potentially gendered lifestyle variables were not recorded or were unavailable. Pregnant and breastfeeding females, females who had undergone a hysterectomy, and individuals undergoing treatment with sex hormones other than OCs were not evaluated in this investigation. Molecular alterations in these cases should be considered in future and included in the classifier. These subjects comprised approximately 5% of the samples in this study, but could be important for investigating other populations. Furthermore, the links between these sex and hormonal status variations in serum and brain analyte concentrations and brain function should be investigated. Future studies should further examine sex-specific mechanisms in inflammatory, metabolic, and cardiovascular processes in peripheral and central tissues and in control and patient populations. This may shed more light on sex-based differences in mental disorders.

This investigation has established sex differences in serum concentrations of a number of proteins and small molecules involved in endocrine, immune, metabolic, and other biological functions. These are largely consistent with the *Ramsey et al (2012)* (*341*) meta-analysis and a number of novel findings in this investigation have provided further contributions to the body of literature on sex differences. Many of the investigated analytes were found to vary with female hormonal status and have previously been associated with schizophrenia and MDD. These results provide insight into the influence of sex and hormonal variation on biological processes and their potential impact on sex differences in the mechanisms involved in mental disorders. However, the relationships between brain and peripheral effects of these analytes have not been established. The extensive differences between males and females with menstrual cycles, using OCs, and after menopause also pose a challenge of previously undocumented significance to biomarker studies. A high rate of false

positive results was observed in groups with simulated imbalances in the proportions of males or OC users. Therefore, a classifier was constructed capable of grouping subjects by sex and female hormonal status. Future studies should be designed to collect this information, but this classifier may be useful to investigators wishing to retrospectively evaluate the effects of sex and hormonal status on biomarker studies.

Chapter 4 Sex differences in serum markers of major depressive disorder in the Netherlands Study of Depression and Anxiety

4.1 Introduction

Major depressive disorder (MDD) is a highly prevalent, disabling, and recurrent condition. A higher prevalence of MDD in women (approximately twice that of males) has been consistently found in many countries and cultures beginning in late adolescence (*150*). Comorbid anxiety disorders are present in more than half of those suffering from MDD and are also more prevalent in women (*155*). Studies have suggested that sex differences in biological processes may contribute to the increased risk of MDD in females. Hypotheses implicating hypothalamic-pituitary-adrenal (HPA) (*161*) and hypothalamic-pituitary-gonadal (HPG) axis dysfunction (*160*, *164*), inflammation and immune function, and others involving genetic factors, the hypothalamic-pituitary-thyroid (HPT) axis, and neurotransmitter and brain function (*156*, *157*) have emerged (see **Chapter 1**). However, no convincing explanation for the higher rate of MDD in females has been found.

Furthermore, changes in peripheral inflammation, oxidative stress, and metabolic markers, growth factors, and endocrine factors have been associated with MDD and these have been proposed as potential biomarkers (*122*, *369*). Sex differences in these pathways have been demonstrated in **Chapter 3** and elsewhere (see **Chapter 1**). Biomarkers may help to diagnose individuals as part of a screening programme (*369*) or to improve recognition and treatment of MDD in primary care settings (*91*, *370*). Recently, differences between MDD patients and controls have been found in their blood metabolic, transcriptomic, and proteomic profiles (*129*–*131*, *371*, *372*), although further validation of these is required. Sex differences in biomarkers of MDD were only reported in the work of *Domenici et al (2010)* (*130*), who found 11 plasma analytes with significant interactions between sex and diagnosis, including growth hormone and proteins involved in immune response. However, these were not evaluated as predictive markers and subjects were medicated MDD patients with recurrent episodes, who were not extensively characterized for psychiatric comorbidities, demographic, lifestyle, or health characteristics, and who were not followed up. The importance of sex-dependent markers of MDD requires further investigation, as highlighted by a recent finding that elevated morning cortisol was a biomarker for later

depression only in adolescent boys with high depressive symptoms (*166*). An extensive investigation of sex differences in a large panel of potential biomarkers of MDD may help to elucidate sex differences in the pathophysiology of MDD in addition to their potential use in diagnosis.

4.1.1 Aims and objectives

The investigations carried out in the present study had the following aims and objectives:

1.) To identify sex differences in serum biomarkers and potential pathophysiological mechanisms in MDD using a Human DiscoveryMAP® multiplex immunoassay panel of 243 molecules. These were then also evaluated in patients with comorbid MDD and anxiety disorder(s) (CMA) and in remitted individuals previously diagnosed with MDD and not currently experiencing an episode.

2.) To evaluate the performance of a panel of analytes for MDD/control classification using logistic regression for males and females separately. The performance of these sex-specific models was then tested in CMA and remitted MDD to evaluate specificity of these models for MDD.

3.) To investigate HPG axis function in females by testing how the levels of reproductive hormones vary in the follicular and luteal phases of the menstrual cycle, during oral contraceptive (OC) use, and after menopause in MDD, CMA, and remitted MDD compared to controls. Hormones evaluated were follicle stimulating hormone (FSH), luteinizing hormone (LH), testosterone, and progesterone.

Samples from the Netherlands Study of Depression and Anxiety (NESDA) were used to carry out these investigations. The same multiplex immunoassay panel was used to evaluate overall markers of current and remitted MDD in the NESDA cohort previously *(Bot et al (2015)* (*373*)*)* and to investigate variations in serum analyte concentrations with sex and female hormonal status in NESDA control subjects in **Chapter 3**. *Bot et al (2015)* (*373*) found 28 analytes with significantly different serum concentrations in MDD compared to controls involved in cell communication and signal transduction, protein metabolism, and immune response. These findings included reduced levels of reproductive hormones FSH and LH in MDD compared to controls.

4.2 Methods

4.2.1 Clinical Samples

Serum samples were from NESDA, a large and ongoing cohort study described in **Chapter 3**. Diagnoses of depressive disorders (MDD and dysthymia) and anxiety disorders (social phobia, generalized anxiety disorder, panic disorder, and agoraphobia) were carried out using the Composite Interview Diagnostic Instrument (CIDI) in accordance with Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV criteria (*22*). Patients with current disorders were classified as having a depressive and/or anxiety episode within the past six months and remitted patients had a lifetime diagnosis, but were not currently experiencing an episode. History of MDD and anxiety disorders, age of onset, and presence of first onset or recurrent MDD was determined with the CIDI. Family history of depression or anxiety in first degree relatives was assessed (*374*), as well as antidepressant and benzodiazepine use. Control subjects had neither a current nor a lifetime diagnosis of assessed psychiatric disorders according to the CIDI and did not develop any such disorder by the second year follow-up assessment (N=365). Patients and controls were excluded when not fluent in the Dutch language and when they had a primary clinical diagnosis of other psychiatric disorders not studied in NESDA: bipolar disorder, obsessive compulsive disorder, severe substance use disorder, or psychotic disorder. Further details of the NESDA study design and protocol, including other demographic, lifestyle, and health information assessed, can be found in **Chapter 3** and have also been previously described (*342*).

Patients were categorized into three separate groups. These consisted of MDD patients (current episode; N=231), comorbid MDD and anxiety disorder(s) patients (CMA; concurrent episode of at least one anxiety disorder; N=360), and remitted MDD patients (lifetime diagnosis, but no current episode; N=287). Demographic and other information for the three patient groups (MDD, CMA, and remitted MDD) and the control group is reported in **Table 4.1A-B.**

Blood samples from NESDA participants were collected at the baseline assessment at approximately 0800 hours. As discussed in **Chapter 3,** the Human DiscoveryMAP® multiplex immunoassay platform was used to measure the serum concentrations of 243 total analytes across 1,840 available NESDA participant samples assigned to 26 plates. Three samples were removed after internal quality control checks.

Table 4.1. Female (A) and male (B) demographic, health, and lifestyle characteristics for MDD, CMA, remitted MDD, and controls. Values are shown as mean ± the standard deviation. Differences between controls and individual conditions were assessed using Welch's *t*-test (continuous data) or Fisher's exact test (discrete data). Variables in bold were significantly different ($p < 0.05$) from controls. **Abbreviations:** MDD (major depressive disorder); CMA (comorbid MDD and anxiety disorder(s)); BMI (body mass index); MET (metabolic equivalent of task); OC (oral contraceptive); TCA (tricyclic antidepressant); SSRI (selective serotonin reuptake inhibitor).

(A)	Control (N=225)	MDD (N=149)	CMA (N=259)	Remitted MDD (N=205)
Collection sites				
Area % (Amsterdam/Groningen/Leiden)	18/44/37	21/38/42	19/36/46	16/54/30
Frame % (General population/Primary care/Specialised mental health care)	26/74/0	**11/40/50**	**7/38/55**	**38/62/0**
Lifestyle and demographic variables				
Hormone status (Follicular/Luteal phase/OC use/ Postmenopausal/Other)	34/37/79/57/18	21/35/46/36/11	45/64/71/57/22	31/33/50/69/22
Age (years)	39±14	39.3±13	40.2±12	**43.8±13**
BMI (kg/m^2)	24.6±5	24.9±5	**26.3±6**	**25.8±5**
Waist circumference (cm)	83.3±13	84.3±12	**87.5±16**	**86±12**
North European ancestry (Yes/No)	97/3	95/5	**92/8**	99/1
Education (years)	12.8±3	12.4±3	**11.1±3**	12.5±3
Partner % (Yes/No)	78/22	**62/38**	**65/35**	74/26
Smoking % (Never/ex-smoker/not regular/regular)	44/34/10/12	**32/30/13/25**	**29/29/13/29**	**32/36/14/19**
Alcohol % (drinks per week)	5±6	5.2±7	5.1±10	6±7
Drug use in last month % (Yes/No/NA)	4/96/0	3/97/0	5/93/2	6/94/0
Physical activity (MET)	3819.4±2893	3436.6±3112	3448.4±3028	4107.2±3137
Health and medication				
Systolic blood pressure (mm Hg)	129.4±20	129.2±17	132.8±21	132.8±21
Diastolic blood pressure (mm Hg)	77.6±11	79.4±9	**81.6±13**	**80.5±11**
Chronic disease % (Yes/No)	31/69	**44/56**	**46/54**	39/61
Anti-inflammatory medication % (Yes/No)	3/97	6/94	6/94	7/93
Lipid modifying agents % (Yes/No)	3/97	1/99	5/95	6/94
Antihypertensive medication % (Yes/No)	10/90	8/92	12/88	15/85
Depression characteristics				
Symptoms (Inventory of Depressive Symptomatology; IDS)	8.6±7	**26.6±11**	**34.2±12**	**14.9±9**
Symptoms (Becks Anxiety Inventory; BAI)	4.2±5	**11.5±7**	**20.1±11**	**5.9±6**
Family history % (Yes/No)	72/28	**83/17**	**86/14**	**83/17**
MDD type % (first episode/recurrent)		43/57	43/57	54/46
Anxiety % (Lifetime/No lifetime diagnosis)		34/66	100/0	47/53
Benzodiazepine use % (Yes/No)		11/89	18/82	1/99
Antidepressant use % (TCA/SSRI/Other/Mixed/None)		1/28/9/1/61	5/32/7/1/55	1/13/1/85

(B)	Control (N=140)	MDD (N=82)	CMA (N=101)	Remitted MDD (N=82)
Collection sites				
Area % (Amsterdam/Groningen/Leiden)	11/51/38	13/50/37	17/39/45	15/48/38
Frame % (General population/Primary care/Specialised mental health care)	33/67/0	**7/35/57**	**6/33/61**	**57/43/0**
Lifestyle and demographic variables				
Age (years)	39.7±15	**44.8±12**	**43.8±11**	42.9±14
BMI (kg/m^2)	25.3±4	26.1±5	**26.5±5**	26.3±4
Waist circumference (cm)	93.1±13	96.1±14	**97.4±14**	95.8±12
North European ancestry (Yes/No)	99/1	99/1	97/3	96/4
Education (years)	12.7±3	12±3	**11.3±3**	11.8±3
Partner % (Yes/No)	74/26	76/24	71/29	78/22
Smoking % (Never/ex-smoker/not regular/regular)	32/37/14/16	21/38/15/27	27/33/10/31	**22/28/22/28**
Alcohol (drinks per week)	10.8±11	11±15	11.6±15	10.3±12
Drug use in last month % (Yes/No/NA)	9/89/2	9/91/0	9/90/1	6/93/1
Physical activity (MET)	4154.5±3551	3521.2±3118	**3291.3±2668**	4317.8±3530
Health and medication				
Systolic blood pressure (mm Hg)	144.8±18	142.5±18	143.8±26	143.9±18
Diastolic blood pressure (mm Hg)	81.7±12	83.9±10	85.2±15	83.8±11
Chronic disease % (Yes/No)	31/69	**45/55**	**54/46**	**45/55**
Anti-inflammatory medication % (Yes/No)	1/99	**9/91**	7/93	5/95
Lipid modifying agents % (Yes/No)	8/92	**20/80**	11/89	9/91
Antihypertensive medication % (Yes/No)	15/85	21/79	15/85	16/84
Depression characteristics				
Symptoms (Inventory of Depressive Symptomatology; IDS)	5.5±5	**27.6±11**	**34.3±13**	**13.8±9**
Symptoms (Becks Anxiety Inventory; BAI)	2.5±4	**11.9±9**	**20.8±12**	**4.1±5**
Family history % (Yes/No)	66/34	**82/18**	**80/20**	78/22
MDD type % (first episode/recurrent)		40/60	52/48	52/48
Anxiety % (Lifetime/No lifetime diagnosis)		22/78	100/0	46/54
Benzodiazepine use % (Yes/No)		6/94	16/84	2/98
Antidepressant use % (TCA/SSRI/Other/Mixed/None)		4/16/6/2/72	3/32/16/0/50	1/10/0/89

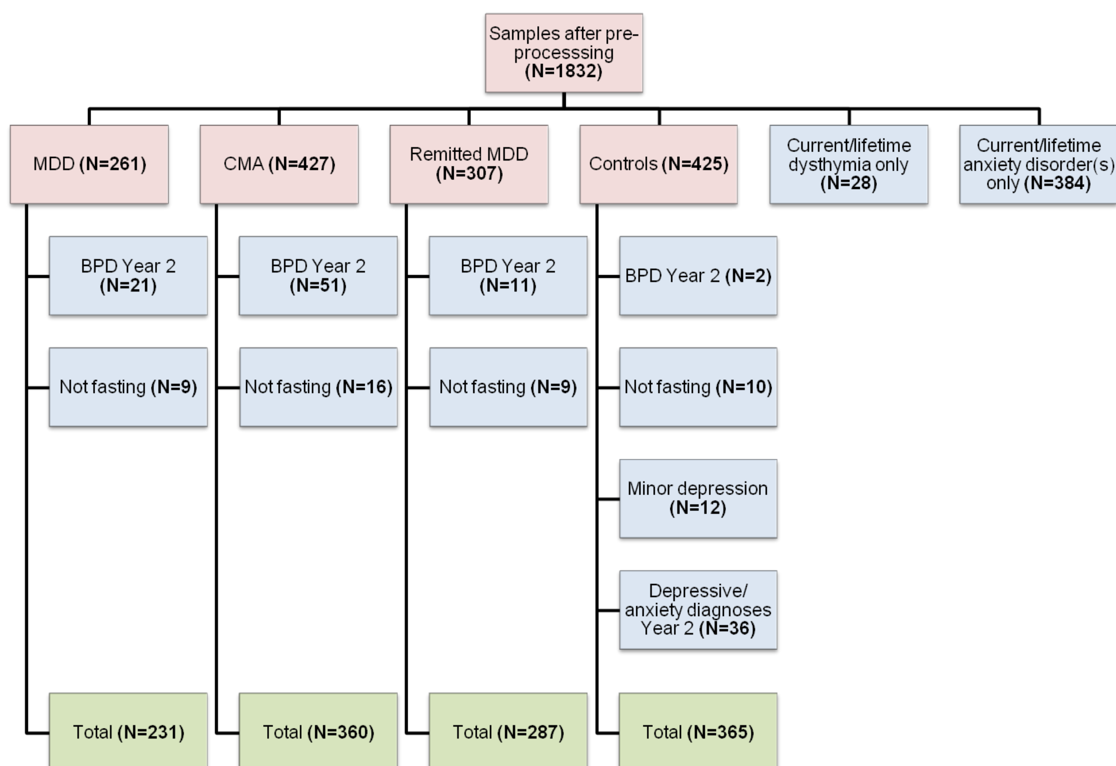
4.2.2 Data pre-processing

Data analysis and pre-processing was carried out as described in **Chapters 2** and **3.** Seventy-two assays and one sample were removed with more than 30% missing values. ComBat was used after $\log_2$ transforming the analyte data to adjust for plate effects and multivariate outlier assessment (**Chapter 2**) resulted in the removal of four samples. Missing covariate and analyte data were replaced as described in **Chapter 2**.

4.2.3 Sample exclusion

A total of 1,832 samples remained after data pre-processing. Additional samples were excluded from analysis in this study for the following reasons, as shown in **Figure 4.1**: 1) the participants had not fasted when blood was withdrawn (N=44); or the participants were diagnosed with 2) current or lifetime dysthymia only (N=28), 3) current or lifetime anxiety disorder(s) only (N=384), 4) current minor depression only (N=12), 5) bipolar disorder (lifetime diagnosis) in the second year follow-up assessment (N=85), or 6) depressive and/or anxiety disorder episode within two years of the baseline measurement (N=36). This left 1,243 samples for further analysis.

Figure 4.1. Sample exclusion for patient groups and controls after data pre-processing. Numbers are indicated in brackets. Red boxes = initial data after pre-processing; blue boxes = excluded data; green boxes = data after exclusion. Participants were excluded if they were not fasting or if they had a current/lifetime dysthymia or anxiety disorder only, current minor depression only, a depressive and/or anxiety diagnosis within two years of the baseline measurement, or a lifetime bipolar disorder diagnosed at the two-year assessment.



4.2.4 Data analysis

A logistic regression model was used with patient/control status as the outcome variable to test for sex differences in the associations between log_2-transformed analyte concentration and log-odds of MDD diagnosis, indicated by a significant interaction between analyte level and sex. Logistic regression and interactions are further described in **Chapter 2.** Sex, analyte concentration, and their interaction were included in the model. Additional covariates (further described in **Chapter 3** and summarized in **Table 4.1A-B**) were also considered, including: ancestry, education, diastolic blood pressure, physical activity, family history of anxiety or depression, presence of chronic disease, use of lipid modifying agents, use of anti-inflammatory drugs, use of antihypertensive medication, collection area, BMI, age, alcohol consumption, smoking status, recreational drug use, partner status, and hormonal status (follicular phase of menstrual cycle/luteal phase/use of oral contraceptives (OCs)/postmenopausal status/other), as well as interactions between them and sex. These additional covariates were selected using stepwise regression with simultaneous forward selection and backward elimination using the Bayesian Information Criterion (BIC) selection criterion. Analytes with significant sex-analyte interaction terms ($p<0.05$) were tested separately in males and females for each disorder and classified as either male-specific, female-specific or qualitative markers. Male-specific markers were significant in males only ($p<0.05$ in males, $p>0.20$ in females) and female-specific markers were significant in females only ($p<0.05$ in females, $p>0.20$ in males). Qualitative markers had opposite associations with diagnosis between males and females. The false discovery rate (FDR) analogue of the *p*-value (*q*-value) was calculated for all analyses, described in **Chapter 2**. Ratios of the geometric means of analyte concentrations between patient and control groups were calculated after adjustment for batch effects using ComBat. MDD/control sex interaction tests using log_2-transformed analyte concentrations were compared with analyses using untransformed analyte concentrations. Significant analytes from log_2-transformed analyses were subsequently tested for sex-analyte interactions in CMA and in remitted MDD compared to controls. Note that the same control subjects were used as the reference population for MDD, CMA, and remitted MDD. Significant differences between the sex-dependent markers of MDD and CMA were also tested, using CMA/MDD status as the outcome variable and checking for significant interactions between log_2-transformed analyte level and diagnosis. Variation in sex-analyte interactions with female hormonal status (follicular phase of the menstrual cycle, luteal phase, OC use, and postmenopausal status, other) were evaluated by testing for significant interactions between log_2-transformed analyte concentration and female hormonal status. Gene ontology (GO) biological process (BP) terms for significant proteins were assessed using the UniProt website (*www.uniprot.org;*

accessed March 2015) and European Bioinformatics Institute's QuickGO online database *(www.ebi.ac.uk/QuickGO; accessed March 2015)*.

4.2.5 Joint analyte effects and classification

To evaluate the predictive performance of a set of molecular markers in classification of male and female MDD, a forward stepwise logistic regression model with BIC was used to define the optimal set from all 171 analytes measured and retained after quality control. Details on assessment of classification performance can be found in **Chapter 2** (**Section 2.3.13**). Briefly, a repeated ten-fold cross-validation procedure was used to estimate accuracy of the models for males and females separately. In this procedure, ten-fold cross-validation was repeated 50 times in order to allow for random partitioning of the data in each cross-validation step. Average receiver operating characteristic (ROC) curves were then plotted and area under the ROC curve (AUC) was calculated. Finally, the specificity of the model for MDD was evaluated by testing its ability to correctly classify CMA and remitted MDD.

4.2.6 Female HPG axis function

Generalized least squares (GLS) regression was used to assess female HPG axis function by evaluating the variation in $\log_2$-transformed levels of reproductive hormones in controls compared to MDD, CMA, and remitted MDD conditions in females in the follicular and luteal phases of the menstrual cycle, using OCs, and after menopause. GLS was used to account for heteroskedasticity in hormone levels across hormonal status and diagnosis and was implemented using the nlme package (*320*) in R. Main effects for diagnosis and interactions between hormonal status and diagnosis were evaluated for testosterone, progesterone, follicle stimulating hormone (FSH) and luteinizing hormone (LH). Where testosterone showed a significant main effect or interaction, sex hormone-binding globulin (SHBG) was also evaluated to determine how levels of free testosterone may be affected. Additional covariates (further described in **Chapter 3** and summarized in **Table 4.1A-B**) were again selected using stepwise regression with simultaneous forward selection and backward elimination using the BIC selection criterion. *P*-values less than 0.05 were considered statistically significant.

4.3 Results

In the NESDA cohort, females were overrepresented in the MDD (65%), CMA (72%), remitted MDD (71%), and control (62%) classifications (**Table 4.1A-B**). MDD, CMA, and remitted patients had significantly higher diastolic blood pressure and were more likely than

controls to suffer from a chronic disease, use anti-inflammatory medication, smoke, and have a family history of depression or anxiety. CMA and remitted patients were older and had a higher BMI, while CMA and MDD patients were less likely to have a partner. CMA patients also had higher a waist circumference, fewer years of education, were less physically active, and were less likely to have North European ancestry. There were no significant sex differences in MDD type (first episode or recurrent), benzodiazepine or antidepressant use, IDS or BAI scores, or presence of lifetime anxiety disorder diagnosis in MDD, CMA, or remitted MDD.

Twenty-eight analytes had a significant interaction between log_2-transformed serum concentration and sex in the logistic regression analysis of MDD patients and controls (**Table 4.2**; **Figure 4.2**). Presence of family history of depression or anxiety, use of anti-inflammatory medication, partner status, age, sex-age interaction, and chronic disease were selected as additional covariates to sex, analyte level, and their interaction in the stepwise logistic regression analyses. Insulin-like growth factor binding protein (IGFBP)-5, macrophage inflammatory protein (MIP)-3β, and interleukin (IL)-2 receptor α were the only female-specific markers, with lower serum levels associated with increased odds of MDD in females. Of the remaining 25 analytes, 12 were male-specific and 13 had a qualitative interaction (males and females had opposite associations between MDD diagnosis and serum concentration). With the exception of hepatocyte growth factor (HGF) receptor and eotaxin-1, higher levels of male-specific analytes and analytes with qualitative interactions were associated with increased odds of MDD relative to females. Eight of the analytes [trefoil factor 3 (TFF3), C-reactive protein (CRP), CD5 antigen-like (CD5L), IGFBP-4, β2-microglobulin (B2M), osteoprotegerin (OPG), tumor necrosis factor receptor 2 (TNFR2), and fetuin-A] were also significant at $q<0.10$. Many of the significant analytes were involved in defence, inflammatory, and/or immune response. Others, including IGFBP-4 and -5, HGF receptor, von Willebrand factor (vWF), urokinase-type plasminogen activator receptor (uPAR), thyroxine binding globulin (TBG), pancreatic polypeptide (PPP), tenascin-C (TN-C), OPG, factor VII, and myoglobin, participated in a variety of biological processes. Among these were blood coagulation, apoptosis, signal transduction, cell adhesion, oxygen transport, regulation of cell growth, glucose metabolism, skeletal system development, and regulation of appetite.

Most results remained significant using untransformed serum concentrations compared to log_2-transformed concentrations (**Table 4.2**; **Figure 4.2**). However, interactions between sex and IGFBP-5, FASLG receptor (FAS), TN-C, PPP, and uPAR were no longer significant using untransformed concentrations. Most significant sex-dependent markers of MDD were also not altered with female hormonal status, with only one analyte showing a significant

interaction between female hormonal status and concentration. Reduced levels of uPAR were associated with increased odds of MDD in females in the follicular phase of the menstrual cycle and not in females in the luteal phase, OC users, postmenopausal females, or in other aggregated hormonal statuses (data not shown).

Table 4.2. Analytes with significant interactions between log$_2$-transformed serum concentration and sex in MDD compared to controls. Interaction *p*-values are also shown for untransformed (raw) serum concentrations. Odds ratios (OR) represent the ratio of odds of MDD associated with a two-fold increase in the serum concentration of that analyte from the logistic model. Ratio (R) represents the ratio between the geometric means of patient and control analyte concentrations. Orange boxes = male-specific analytes; white boxes = analytes with a qualitative interaction; blue boxes = female-specific analytes. Presence of significant differences in sex-dependent markers between MDD and CMA are marked with [×]. **Abbreviations:** R (ratio, patient/control); OR (odds ratio); P (*p*-value); Q (*q*-value); CMA (comorbid MDD and anxiety disorder(s)). Analyte abbreviations can be found in the **Appendix** (**Table A.1**).

Analytes	Interaction		Interaction *p*-value (raw)	Females				Males				Different in CMA
	P	Q		OR	P	Q	R	OR	P	Q	R	
TFF3	3.E-07	5.E-05	9E-07	0.95	0.46	0.52	0.90	**7.36**	**5.E-06**	**3.E-04**	1.21	×
IGFBP-4	8.E-04	0.029	0.001	0.72	0.38	0.48	0.98	**5.74**	**8.E-04**	**0.008**	1.12	
B2M	8.E-04	0.029	0.001	0.70	0.27	0.38	0.98	**3.74**	**8.E-04**	**0.008**	1.14	×
Fetuin-A	0.004	0.080	0.002	0.91	0.77	0.80	0.99	**4.50**	**9.E-04**	**0.008**	1.10	
uPAR	0.007	0.12	0.14	0.83	0.28	0.38	1.00	**1.80**	**0.012**	**0.058**	1.20	
Cystatin-C	0.010	0.13	0.013	1.50	0.41	0.49	1.02	**12.91**	**2.E-04**	**0.006**	1.09	×
FABP, adipocyte	0.012	0.15	0.006	1.00	0.98	0.98	1.03	**1.87**	**0.002**	**0.014**	1.26	
TBG	0.018	0.19	0.014	0.74	0.31	0.41	0.97	**2.94**	**0.033**	**0.085**	1.06	×
HGF receptor	0.025	0.23	0.014	1.13	0.68	0.72	1.00	**0.37**	**0.018**	**0.063**	0.90	×
vWF	0.028	0.24	0.020	0.92	0.65	0.69	1.00	**1.72**	**0.019**	**0.063**	1.15	×
PPP	0.038	0.29	0.24	1.02	0.88	0.90	1.05	**1.51**	**0.010**	**0.058**	1.30	
TN-C	0.049	0.30	0.078	0.86	0.55	0.62	0.99	**1.83**	**0.042**	**0.10**	1.07	×
CRP	7.E-05	0.005	9E-07	**0.87**	**0.019**	**0.063**	0.73	**1.29**	**0.002**	**0.014**	1.76	×
CD5L	9.E-05	0.005	7E-05	**0.47**	**8.E-04**	**0.008**	0.89	**1.97**	**0.024**	**0.069**	1.13	×
OPG	0.001	0.032	0.002	**0.55**	**0.030**	**0.080**	0.95	**2.72**	**0.020**	**0.063**	1.09	×
TNFR2	0.001	0.032	7E-04	**0.45**	**0.012**	**0.058**	0.95	**2.27**	**0.043**	**0.10**	1.09	×
VCAM-1	0.007	0.12	0.006	**0.40**	**0.013**	**0.058**	0.96	1.97	0.15	0.23	1.05	×
Factor VII	0.008	0.13	0.015	**0.59**	**0.050**	**0.10**	0.95	1.96	0.078	0.14	1.07	
Myoglobin	0.010	0.13	0.023	0.72	0.084	0.15	0.94	1.59	0.059	0.12	1.09	
FAS	0.015	0.17	0.39	0.78	0.18	0.26	0.97	**1.54**	**0.050**	**0.10**	1.19	
Eotaxin-1	0.019	0.19	0.004	**1.35**	**0.044**	**0.10**	1.13	0.74	0.15	0.23	0.97	×
C-Peptide	0.023	0.22	0.023	0.76	0.13	0.21	0.95	1.53	0.091	0.15	1.11	
MDC	0.035	0.29	0.011	0.71	0.18	0.26	0.97	1.74	0.11	0.18	1.04	
PARC	0.043	0.30	0.035	0.87	0.33	0.43	0.95	1.48	0.082	0.15	1.13	
MMP-7	0.043	0.30	0.022	0.79	0.40	0.49	0.98	1.94	0.067	0.13	1.10	
MIP-3β	0.039	0.29	0.050	**0.61**	**0.019**	**0.063**	0.92	1.32	0.37	0.47	1.03	×
IGFBP-5	0.046	0.30	0.066	**0.29**	**0.013**	**0.058**	0.96	1.42	0.58	0.63	1.00	
IL-2 receptor α	0.049	0.30	0.035	**0.55**	**0.029**	**0.080**	0.94	1.30	0.45	0.52	1.03	×

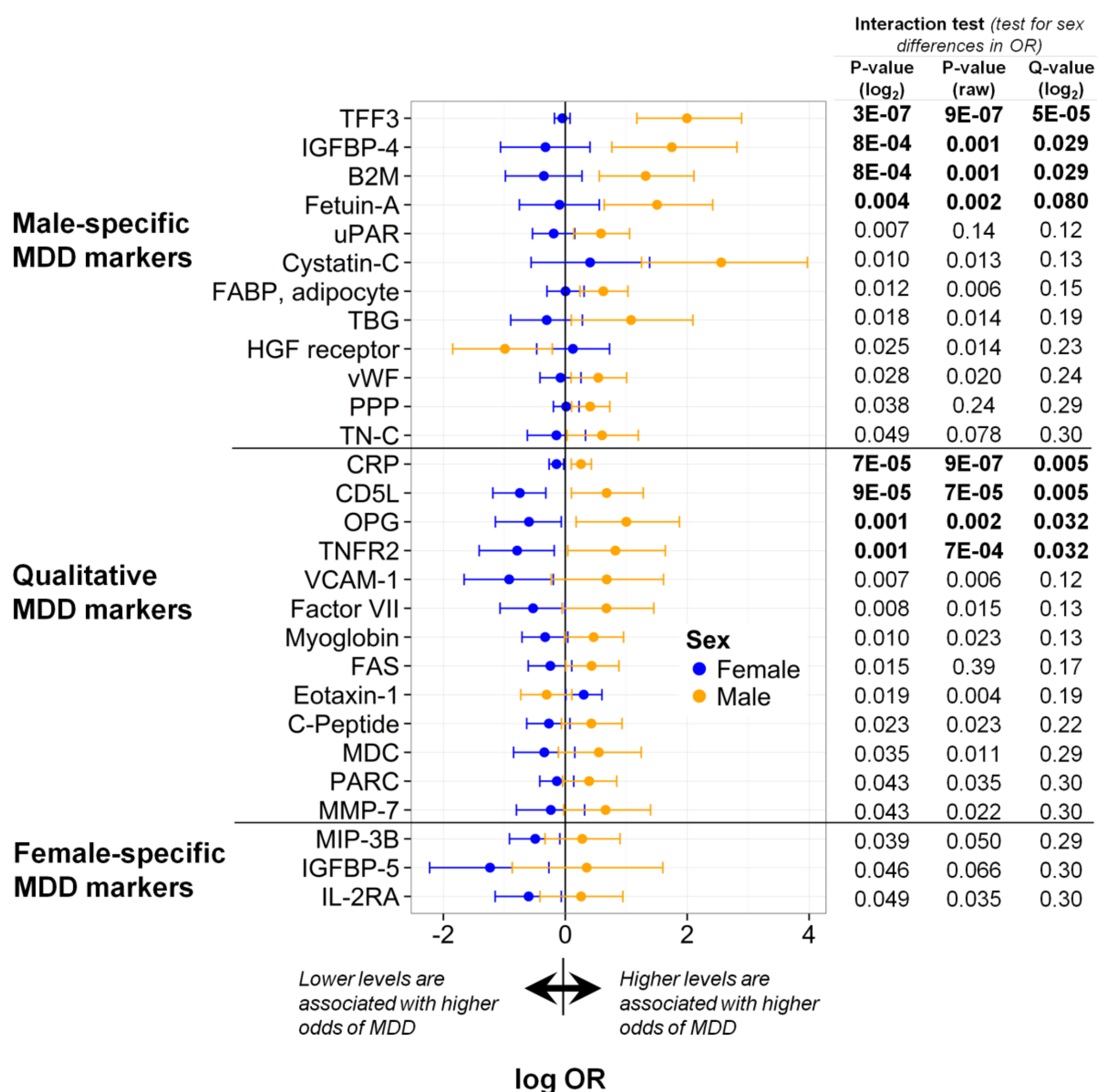


Figure 4.2. Analytes with significant interactions between $\log_2$-transformed serum concentration and sex in MDD compared to controls. Interaction *p*-values are also shown for untransformed (raw) serum concentrations. The natural logarithm of the odds ratio (OR; the ratio of the odds of MDD diagnosis associated with a two-fold increase in the serum concentration of that analyte from the logistic model) is shown for males and females with 95% confidence intervals. **Abbreviations:** OR (odds ratio); P (*p*-value); Q (*q*-value). Analyte abbreviations can be found in the **Appendix** (**Table A.1**).

Five analytes with significant interactions between sex and $\log_2$-transformed analyte concentration in the CMA/control and four in the remitted MDD/control comparisons overlapped with the 28 in the MDD/control comparisons (**Table 4.3**; **Figure 4.3**). Two analytes, TFF3 and factor VII, had significant sex differences in associations with all three conditions. However, most sex-dependent markers of MDD did not overlap with CMA and fifteen had significantly different sex-dependent associations with CMA (**Table 4.2**), including many of the most significant sex-dependent markers of MDD, such as CRP, TFF3, CD5L, TNFR2, osteoprotegerin, and β2-microglobulin.

Table 4.3. Analytes with overlapping significant sex-log$_2$ concentration interactions between MDD and CMA (A) and between MDD and remitted MDD (B) compared to controls. Odds ratios (OR) represent the ratio of odds of the condition associated with a two-fold increase in the serum concentration of that analyte from the logistic model. Ratio (R) represents the ratio between the geometric means of condition and control analyte concentrations. **Abbreviations:** R (ratio, patient/control); OR (odds ratio); P (*p*-value); Q (*q*-value); MDD (major depressive disorder); CMA (comorbid MDD and anxiety disorder(s)). Analyte abbreviations can be found in the **Appendix** (**Table A.1**).

(A)	CMA									
	Interaction		Females				Males			
Analytes	P	Q	OR	P	Q	R	OR	P	Q	R
IGFBP-4	0.002	0.057	0.90	0.74	0.74	1.01	**6.01**	**0.001**	**0.010**	1.09
Myoglobin	0.009	0.11	**0.69**	**0.03**	**0.095**	0.97	1.44	0.12	0.19	1.10
FAS	0.012	0.11	0.80	0.19	0.23	1.01	**1.70**	**0.034**	**0.095**	1.15
TFF3	0.024	0.14	0.92	0.17	0.23	0.80	**2.26**	**0.047**	**0.095**	1.10
Factor VII	0.025	0.14	0.80	0.39	0.43	1.00	**2.18**	**0.041**	**0.095**	1.10
(B)	Remitted MDD									
Factor VII	0.006	0.10	0.74	0.25	0.40	0.97	**2.67**	**0.016**	**0.043**	1.12
TFF3	0.007	0.10	0.95	0.40	0.53	0.78	**3.38**	**0.011**	**0.043**	1.11
MIP-3β	0.039	0.23	0.97	0.87	0.87	1.01	**1.97**	**0.024**	**0.047**	1.11
B2M	0.049	0.23	1.08	0.80	0.87	1.03	**2.89**	**0.013**	**0.043**	1.11

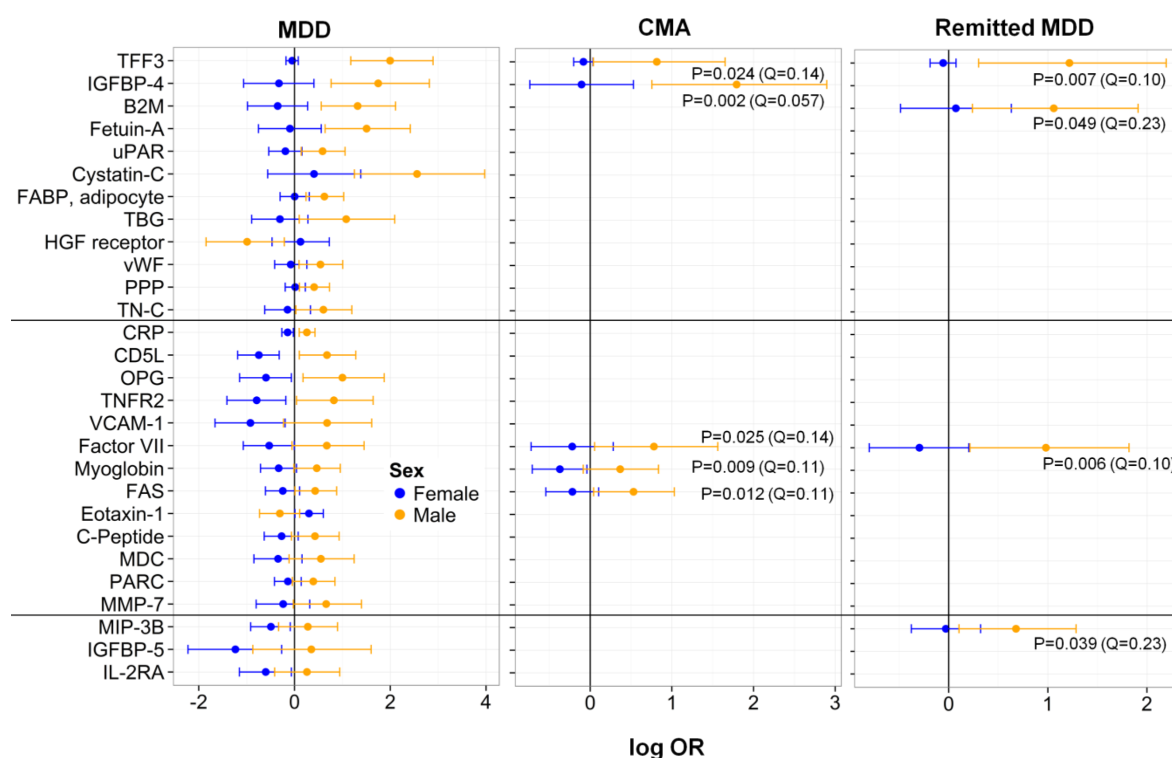


Figure 4.3. Analytes with overlapping significant sex-log$_2$ concentration interactions between MDD, CMA, and remitted MDD compared to controls. The natural logarithm of the odds ratio (OR; ratio of the odds of diagnosis associated with a two-fold increase in the serum concentration of that analyte from the logistic model) is shown for males and females with 95% confidence intervals. *P*- and *q*-values are shown for the interaction tests. **Abbreviations:** OR (odds ratio); P (*p*-value); Q (*q*-value); MDD (major depressive disorder); CMA (comorbid MDD and anxiety disorder(s)). Analyte abbreviations can be found in the **Appendix** (**Table A.1**).

Next, the classification performance of the forward stepwise logistic regression using the 171 measured analytes was evaluated in males and females with MDD separately. Based on repeated ten-fold cross-validation, a AUC of 0.63 was found for males and a AUC of 0.50 was found for females. The ROC curves for males and females and the analytes selected

using forward stepwise logistic regression are shown in **Figure 4.4**. Two analytes were selected for female classification (CD5L and IGFBP5), and five were selected for male classification (TFF3, angiogenin, transthyretin, HGF receptor, and epidermal growth factor (EGF) receptor). Performance of the classifier in males with CMA and remitted MDD compared to controls was also tested. Using the fitted logistic regression model found for males in MDD, AUCs of 0.69 and 0.62, were obtained for male remitted MDD and CMA, respectively.

Serum levels of FSH and testosterone varied with diagnosis in females (**Figure 4.5**) in the linear model. Reduced FSH was found in MDD (regression coefficient $\beta = -0.23$, -14.7% impact of diagnosis on analyte level) and remitted MDD ($\beta = -0.23,\ -14.7\%$), and increased testosterone was found in CMA ($\beta = 0.096,\ 6.8\%$). Variations in serum testosterone across hormonal status also differed between female MDD and controls, with increased levels in OC users with MDD compared to control OC users ($\beta = 0.17,\ 12.6\%$). SHBG levels were not significantly different in CMA (p=0.99) and did not show a significant interaction between diagnosis and hormonal status in MDD (p=0.91), suggesting that differences in total testosterone reflected differences in free testosterone.

Figure 4.4. ROC curve illustrating classification of MDD for males and females using repeated ten-fold cross validation of BIC forward stepwise logistic regression. The logistic regression model for male MDD was also used to classify male CMA and remitted MDD. The legend shows AUCs in brackets. Markers selected in BIC forward stepwise logistic regression of all data for classification of female and male MDD are shown in the right panel. **Abbreviations:** AUC (area under the curve); MDD (major depressive disorder); CMA (comorbid major depressive and anxiety disorder(s)). Analyte abbreviations can be found in the **Appendix** (**Table A.1**).

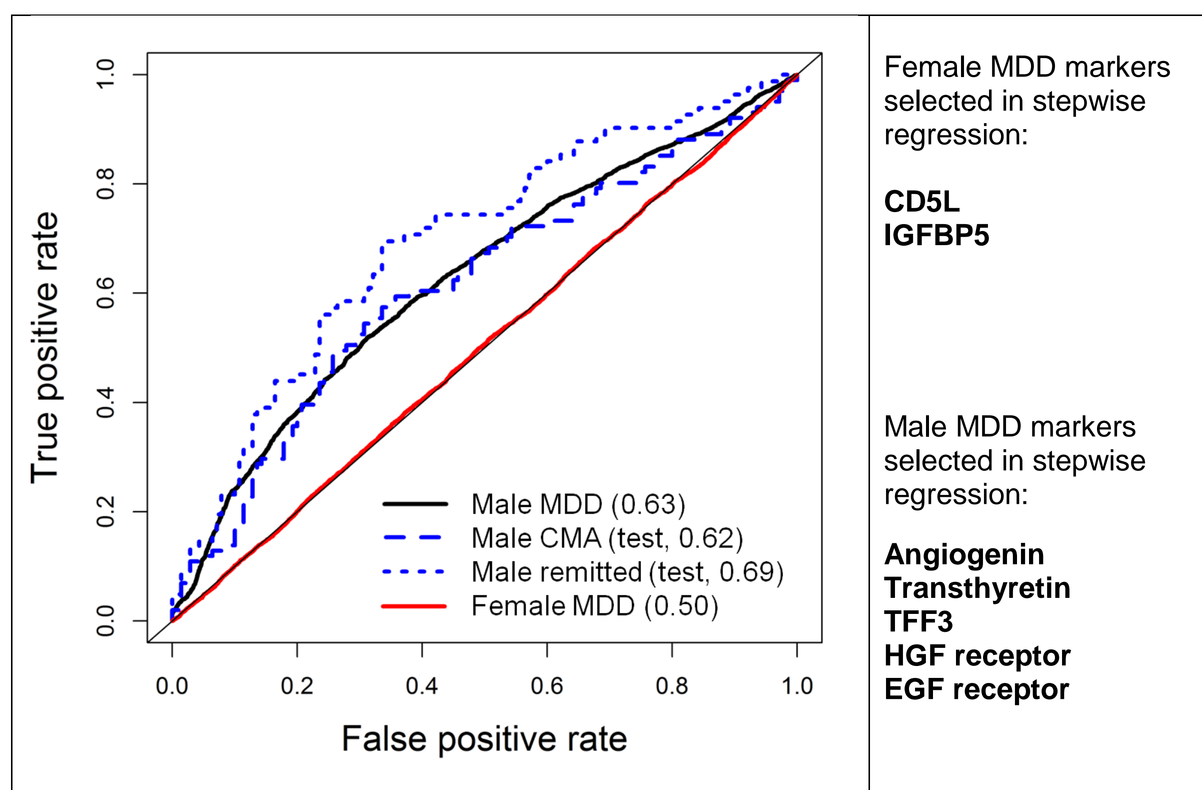
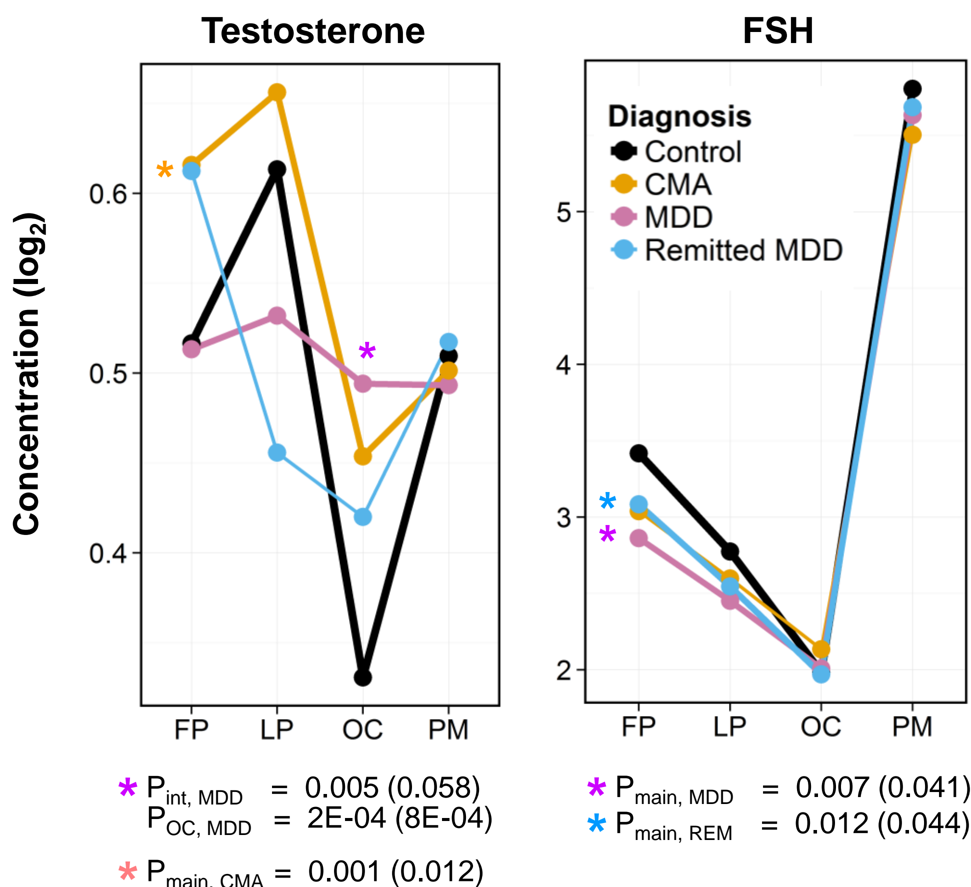


Figure 4.5. Serum concentrations of testosterone and FSH in controls and current and remitted female MDD. Mean log_2-transformed concentrations are plotted for each analyte across follicular and luteal phases of the menstrual cycle, during OC use, and after menopause for female controls, MDD, CMA, and remitted MDD conditions. Where a significant main effect of diagnosis or interaction with hormonal status was found, a star is marked and the *p*-value is written below the plot with the *q*-value in brackets. **Abbreviations:** MDD (major depressive disorder); CMA (comorbid MDD and anxiety disorder(s)); REM (remitted MDD); FSH (follicle stimulating hormone); FP (follicular phase); LP (luteal phase); OC (oral contraceptive use); PM (postmenopausal); int (interaction); P (*p*-value); Q (*q*-value).



4.4 Discussion

This chapter identified sex-dependent serum markers of MDD in a large, well-characterized cohort. Previous studies have measured only a few molecules testing specific hypotheses, assessed limited information about participants, and/or studied specific patient populations. NESDA provides extensive information on a large number of serum analytes and sample characteristics and importantly, includes comorbid disorders and follow-up diagnoses in a large cross-section of the patient population from the community, primary care, and specialized health care settings.

The present study found evidence to support a link between MDD and elevated levels of certain molecules involved in inflammation, immune, and defence response specifically in males, including CRP, TFF3, cystatin-C, fetuin-A, β2-microglobulin, CD5L, and TNFR2. The

association between elevated CRP and male MDD had been found in a previous NESDA and other large cohort studies (*167–171*, *375*). *Bot et al (2015)* (*373*) also found elevated levels of inflammatory molecules α1-antitrypsin, α1-antichymotrypsin, and MIF in MDD patients compared to controls in the combined sample of male and female NESDA subjects. Previous studies suggested changing hormonal levels may be responsible for the absence of association between elevated CRP and female depression (*168*). However, there was no evidence indicating that female hormonal status interacted with CRP or other significant immune markers in determining odds of MDD. Alternatively, stress pathways may influence inflammatory processes differently in males and females, as suggested by *Toker et al* (2005) (*169*). This could occur by three-way interactions between the HPA and HPG axes and the immune system (*29*, *355*, *376*, *377*). Chronic stress may impair glucocorticoid regulation of the immune system in males and result in higher levels of inflammatory markers in depression (*144*, *377*, *378*). However, it has also been proposed that inflammation itself causes the behavioural, neuroendocrine, and neurochemical changes in MDD (*142*, *143*). The results of this study indicate that this causal mechanism could be sex-dependent. This would be consistent with extensive sex differences in immune function, including a higher male susceptibility to infection and proneness to autoimmune diseases characterized by a proinflammatory Th1 immune response (*379*, *380*). However, complex bidirectional relationships between depression and inflammation have been observed and further work will be required to elucidate the mechanisms involved. Other mental disorders are also characterized by male-specific elevations in the serum concentrations of inflammatory molecules, including first-onset antipsychotic naive schizophrenia (see **Chapter 5**) and Asperger syndrome (*10*, *381*).

Few female-specific markers were found in this study. There were also no significant sex-dependent or overall associations between total serum cortisol levels and MDD that would provide an indication of HPA axis dysfunction. However, a previous study of NESDA participants found that individuals with a current or remitted diagnosis of MDD had a higher cortisol awakening response compared to controls, as measured using a series of seven salivary samples over the course of one hour (*137*). Additionally, there were no sex differences in associations between MDD and thyroid stimulating hormone that would have indicated a greater prevalence of subclinical hypothyroidism in female MDD (*156*). Most of the sex-dependent markers of MDD in this work did not overlap with analytes found to have significant sex-diagnosis interactions in *Domenici et al (2010)* (*130*). The exceptions to this were MDC and eotaxin, although the specific eotaxin family member was not recorded. This small overlap may be due to differences in the patients included in these studies, who were largely chronic/recurrent and medicated.

Few sex-dependent markers of MDD overlapped with CMA, and several were also significantly different in CMA. This suggests that different sex-dependent pathophysiological mechanisms could be involved in MDD and CMA and that the conditions should be investigated separately for sex differences. Heterogeneity of anxiety disorders may contribute to these differences and should be further investigated in future. Furthermore, most sex differences in MDD markers were only present during an episode and not in remitted MDD. However, elevated TFF3 was significantly associated with male CMA and remitted MDD and was important in classifying male MDD.

The performance of male and female MDD classification based on serum molecular concentrations in this study was poor. The measured markers classified female MDD patients from controls with an accuracy no greater than chance (AUC = 0.50) and classified male MDD patients from controls with an accuracy slightly higher than chance (AUC=0.63). The male MDD model was able to classify male CMA and remitted MDD with similar accuracies, despite having few overlapping markers with MDD. These classifiers would not be adequate to use as diagnostic tools for depression, but male markers could potentially be used together with clinical data to aid in diagnosis.

There was evidence for a disturbance in levels of reproductive hormones in female MDD, CMA, and remitted MDD, providing further evidence for HPG axis dysfunction in female MDD (*164*). Reproductive hormones and their fluctuations across the female life cycle have previously been hypothesized to play a role in female depression as they are known to interact with the HPA axis, autonomic nervous system response to psychosocial stress, and serotonergic processes (*156*, *160*, *382*). Elevated testosterone was found in CMA and specifically in female OC users with MDD compared to controls. Consistent with this finding, higher rates of depression have been found in polycystic ovary syndrome (PCOS) characterized by high levels of androgens in females (*383*, *384*), elevated negative mood-scale scores have been associated with upper normal free testosterone levels slightly in females (*384*), and increased testosterone has been correlated with depression in pubertal females (*385*). FSH was also reduced in MDD and remitted MDD patients compared to controls. This finding was largely consistent with *Bot et al (2015)* (*373*), but contrasted with previous studies showing normal FSH and LH levels in females with menstrual cycles (*386*) and higher levels in females aged 36-45 years with MDD (*387*). Sources of discrepancy may be the small sample size analysed in *Young et al (2000)* (*386*) and the specific focus on perimenopause in *Harlow et al (2003)* (*387*). Future studies should investigate larger samples of females with more precise reporting of hormonal status and record more measures of reproductive function, including estrogen levels.

Certain limitations and future work should be considered. First, future studies are needed to validate these results and investigate a broader range of biological pathways than is presently covered in the multiplex immunoassay panel. In particular, this may reveal female-specific biomarkers of MDD, largely absent in this study. Second, although the present study analyzed current comorbid MDD and anxiety disorder(s) separately, the number of different anxiety disorders included in the study and the high comorbidity between them prevented further investigation of sex differences in markers of specific anxiety disorders comorbid with MDD. Future studies should further investigate the influence of other psychiatric comorbidities and features of MDD not considered here, such as prior anxiety, chronicity of MDD and dysthymia, atypical and melancholic features, and postpartum depression on markers of depression. Third, integration of this molecular data with brain imaging, psychosocial, genetic, and other relevant biological data would enable investigators to consider a greater number of factors thought to influence the higher prevalence of female MDD and potentially elucidate the role of male-specific associations with elevated immune and inflammation markers. Finally, further work is required to understand the role of female hormonal processes in MDD. In particular, hormonal status information is self-reported and the considerable variability in hormonal levels associated with ovulation, perimenopause, and other reproductive events could not be considered here.

This study presents a significant step forward in understanding molecular sex differences in MDD and has important implications for future studies. Distinct differences between males and females were found in molecular markers of MDD, including male-specific associations between MDD and elevated levels of immune and inflammatory molecules. Some of these were shared in CMA and persisted in remitted MDD, but most were not consistent across conditions. Studies should therefore consider the presence of comorbid anxiety disorder(s) in future studies of sex differences in MDD. Sex differences in markers of MDD had implications for the classification of subjects with MDD. Only male MDD could be classified with an accuracy higher than chance, although this was still low. Furthermore, disturbed serum levels of testosterone and FSH in females with current and remitted MDD conditions offer support for the involvement of HPG axis dysfunction in MDD.

Chapter 5 Sex-dependent serum molecular differences in first onset, antipsychotic naive schizophrenia

5.1 Introduction

Schizophrenia is characterized by a higher male incidence (*12*, *236*) and older female age of onset (*12*, *236*, *239*) with better social course and short to medium term prognosis (*12*, *236*, *243*, *244*). Other sex and gender differences in schizophrenia include fewer affective and more negative symptoms in males (*238*) and better response to treatment with antipsychotic medication in premenopausal females (*243*, *245*, *246*). Hypotheses for the differences between males and females in schizophrenia include a differential predisposition to subtypes of schizophrenia (*273–276*), favourable effects of estrogen (*260*), sex differences in genetic risk (*272*), and neurodevelopment theories (*238*, *277*) (see **Chapter 1**). However, evidence for each of these hypotheses is inconclusive and the causes of sex and gender differences in schizophrenia remain elusive.

Previous studies have largely focused on female-specific disturbances of estrogen and other sex hormones in schizophrenia (*261–263*, *267*) without considering other biological pathways that may be important in the disorder. Disturbed inflammatory and immune (*227*, *388*), metabolic and oxidative stress (*51*, *207*, *208*), endocrine (*209*), and growth factor (*210*, *389*) pathways have been reported in peripheral and brain tissue in schizophrenia (see **Chapter 1**). Many of these are altered by sex and female hormonal status in serum, as demonstrated in **Chapter 3**. Previous studies were also limited by small sample sizes, investigated only patients with a chronic course taking antipsychotic medication, and did not use control subjects. Further investigation of a large panel of relevant proteins and small molecules in serum of antipsychotic naive, first onset schizophrenia patients and controls may provide insights into potential sex differences in the aetiology and pathophysiology of schizophrenia.

5.1.1 Aims and objectives

The investigations carried out in the present study had the following aims and objectives:

1.) To assess potential pathophysiological mechanisms in schizophrenia by investigating overall and sex-dependent differences in the serum concentrations of a panel of 190 analytes between antipsychotic naive, first onset schizophrenia patients and control

subjects. Analytes with sex-dependent differences were further tested for confounding with age at first episode.

2.) To investigate the relationship between serum molecular concentrations and Positive and Negative Syndrome Scale (PANSS) scores of schizophrenia for sex-dependent findings in males and females.

3.) To further examine the potential role of these analytes in schizophrenia by assessing the effects of six weeks of antipsychotic treatment on serum molecular concentrations for sex-dependent findings.

These investigations were carried out using Human DiscoveryMAP® multiplex immunoassay data from serum of four independent German cohorts of 133 first onset, antipsychotic naive patients (79 males; 54 females) and 133 control subjects. Patients and controls were matched based on sex, age, body mass index (BMI), waist circumference, and tobacco and cannabis consumption where this information available. These cohorts have individually been used previously to derive highly sensitive and specific biomarker tests for schizophrenia (*212*, *278*) and to find subgroups of schizophrenia patients with alterations in either immune or growth factor/hormone function in schizophrenia (*389*).

5.2 Materials and Methods

5.2.1 Clinical samples

All patients were suffering from their first episode of illness and were antipsychotic naive at the time of sample collection. Diagnoses were carried out using the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV and clinical tests were performed by psychiatrists under Good Clinical Practice-compliance to minimize variability. Only patients diagnosed with the paranoid subtype of schizophrenia (295.30) were included in the study in order to reduce variability caused by possible heterogeneity of the disorder. Symptom severity was assessed using the PANSS rating system. Clinicians had access to all detailed clinical files including medical histories by proxy and referral letters from their general practitioners. This information, together with that provided by family members, was used to confirm first episode and antipsychotic naive status. Controls were recruited in the same time period from the community by advertising or were chosen from a clinical database of volunteers. Patients and controls with severe substance abuse and other medical conditions such as type 2 diabetes mellitus, hypertension, endocrine abnormalities, or cardiovascular or autoimmune diseases were excluded from the study. Controls with a family history of mental

illness were also excluded from the study. All patients and controls were white European. Collection of samples occurred over a period of approximately ten years from 2000 to 2010. Patient and control characteristics for each of the four cohorts are shown in **Table 5.1**. Cohort 1 was from the University of Cologne, Department of Psychiatry, Germany; Cohort 2 was from the University of Muenster, Germany; and Cohorts 3 and 4 were from the University of Magdeburg, Germany. For each cohort, male and female patients were matched to controls according to age, BMI, waist circumference, and tobacco and cannabis consumption where this information was available. Thirty-three patients from Cohort 3 were also assessed before and after a six-week inpatient treatment period (**Table 5.2**). Patients were treated with olanzapine, quetiapine, risperidone, and other atypical and typical antipsychotics. Cumulative chlorpromazine units over the course of treatment were calculated.

The ethical committees of the participating hospitals at the Universities of Cologne, Muenster and Magdeburg each approved the protocols of the study, informed written consent was given by all participants, and studies were conducted according to the Declaration of Helsinki. In cases where participants had a compromised capacity to give consent, next of kin, care takers, or guardians consented on their behalf. Medication was not withheld and was given when possible after diagnosis was completed.

Samples were collected between 0800 and 1000 hours from all participants, with no systematic differences between patients and controls. The Human DiscoveryMAP® multiplex immunoassay platform was used to measure the serum concentrations of 190 total analytes across all samples in the separate cohorts.

5.2.2 Data pre-processing

Data analysis and pre-processing was carried out as described in **Chapter 2**. Forty-seven assays not measured in all cohorts were removed and a further 48 were removed with 30% or more missing values in at least one cohort. A list of the 95 remaining assays and the percentage of missing values contained in each assay can be found in the **Appendix** (**Table A.1**). Remaining missing values were replaced as described in **Chapter 2** and values were $\log_2$-transformed. Outlying samples were assessed in each cohort based on a robust measure of the Mahalanobis distance (**Chapter 2**).

Table 5.1. Patient and control characteristics by cohort. Values are shown as mean ± sd. Tobacco and cannabis use is displayed as (yes/no/NA) and age is in years. Female hormonal status was predicted (see **Chapter 3**) and is displayed as (oral contraceptive user/menstrual cycle/postmenopausal). The final two columns show that there were no significant (p<0.05) differences in age, BMI, smoking, cannabis use, or predicted hormonal status between schizophrenia patients and controls for males or females in the final combined cohort. Welch's t-test was used to test for differences in continuous data and Fisher's exact test was used for categorical data. **Abbreviations:** M (male); F (female); N (number); BMI (body mass index (kg/m^2)); PANSS (Positive and Negative Syndrome Scale); P (positive); N (negative); G (general).

	Cohort 1		**Cohort 2**		**Cohort 3**		**Cohort 4**		**P-values**	
Centre	Cologne		Muenster		Magdeburg		Magdeburg			
Collection period	2000-2009		2000-2007		2008-2010		2010			
	M	F	M	F	M	F	M	F	M	F
Patients N	23	25	34	11	14	10	8	8	-	-
Controls N	23	25	34	11	14	10	8	8	-	-
Patients age	26.4±6	30.4±9	27.6±9	23.5±7	28.5±9	36.2±14	33.3±11	37.0±12	0.81	0.75
Controls age	27.1±6	30.2±7	27.6±8	23.6±7	29.6±9	35.6±14	32.0±10	34.3±10		
Patients BMI	22.8±2	23.1±5	22.7±2	20.9±2	23.7±4	25.3±5	22.2±2	21.2±3	0.55	0.67
Controls BMI	22.7±2	22.3±4			24.5±5	24.9±4	23.2±1	22.0±4		
Patients tobacco	13/7/3	8/8/9	15/15/4	1/10/0	9/5/0	7/3/0	4/4/0	2/5/1	0.13	0.17
Controls tobacco	11/12/0	11/14/0			6/8/0	2/8/0	3/5/0	2/6/0		
Patients cannabis	12/9/2	13/6/6	14/16/4	1/10/0	0/14/0	0/10/0	0/8/0	0/7/1	1.0	0.63
Controls cannabis	13/8/2	12/12/1			0/14/0	0/10/0	0/8/0	0/8/0		
Patients hormonal status (predicted)		7/15/3		6/4/1		3/4/3		1/5/2		0.24
Controls hormonal status (predicted)		11/12/2		7/3/1		3/5/2		4/4/0		
PANSS P1-P7	23.4±5	23.3±5	17.2±6	18.0±10	22.4±7	18.8±6	14.4±5	23.9±9	-	-
PANSS N1-N7	24.5±9	22.7±6	17.7±7	17.4±9	20.1±9	18.3±8	14.9±2	16.4±5	-	-
PANSS G1-G16	49.0±9	50.0±10	37.6±10	36.6±13	44.3±14	40.1±9	33.4±6	40.8±15		
Total PANSS	96.9±19	96.0±19	72.5±20	72.0±30	86.9±27	77.2±15	62.6±11	81.0±25	-	-

Table 5.2. Treated schizophrenia cohort characteristics. First episode patients were from Cohort 3 and were initially antipsychotic naïve. Values are shown as mean ± sd. Tobacco and cannabis use is displayed as (yes/no/NA) and age is in years. **Abbreviations:** M (male); F (female); N (number); BMI (body mass index (kg/m^2)); PANSS (Positive and Negative Syndrome Scale); P (positive); N (negative); G (general); T0 (initial time point); T6 (after six weeks of antipsychotic treatment).

	Antipsychotic follow-up	
Patients N	33	
Patients sex (M/F)	22/11	
Patients age	31.0±10	
Patients BMI	24±4	
Patients tobacco	24/9/0	
Patients cannabis	9/24/0	
	T0	**T6**
PANSS P1-P7	21.3±6	11.6±4
PANSS N1-N7	18.7±9	14.3±6
PANSS G1-G16	42.6±12	28.1±8
Total PANSS	82.6±22	54.0±17
Antipsychotic type (olanzapine/ quetiapine/ risperidone/other)	0	8/7/9/9
Cumulative chlorpromazine units (mg)	0	16 650±14 000

5.2.3 Hormonal status

Hormonal status was not recorded for female subjects in the study. Therefore, the random forest classifier developed in **Chapter 3** was used to classify females as oral contraceptive (OC) users, as having menstrual cycles, or as postmenopausal. Prior to prediction, overlapping analyte data consisting of 79 assays from NESDA and from the current study were combined using ComBat, implemented in the sva package (*303*, *304*) in R.

5.2.4 Data analysis

The data for the combined cohorts were initially analysed to identify analytes present at significantly different levels in schizophrenia compared to control subjects. A stepwise regression was performed with simultaneous forward selection and backward elimination using Bayesian Information Criterion (BIC) to select additional covariates from age, sex, cohort, and interactions between cohort and these variables. Regression diagnostic plots were inspected for analytes with regression outliers, identified with the Bonferoni outlier test for Studentized residuals ($p<0.05$) implemented in the car package (*316*) in R. Robust regression (MM-estimator) was then used, implemented in the robustbase package (*317*) in R. The percentage impact of schizophrenia diagnosis on analyte level was calculated from the regression coefficient. A *p*-value cut-off of 0.05 was used to determine statistical significance.

Next, sex-diagnosis interactions were calculated for each log_2-transformed analyte using stepwise regression with the same additional covariates, implementing robust regression when regression outliers were identified. Analytes found to have significant sex-diagnosis interactions ($p<0.05$) were re-analysed separately in males and females. Analytes were then classified as

male-specific, female-specific, or as having a qualitative or quantitative interaction. Male-specific analytes were significant in males only (*p*<0.05 in males, *p*>0.20 in females) and female-specific analytes were significantly different in females only (*p*<0.05 in females, *p*>0.20 in males). Analytes with qualitative interactions showed opposing differences between males and females, while quantitative interactions showed differences in the same direction, but these differences were greater in one sex. Interaction tests for $\log_2$-transformed analyte concentrations were compared with analyses using untransformed analyte concentrations. Free testosterone levels were also assessed, approximated using the free androgen index (FAI) as the ratio of total testosterone divided by sex hormone binding globulin (SHBG) (*390*). The analysis was then repeated to find $\log_2$-transformed analytes with significant age-diagnosis interactions and potential confounding with sex-diagnosis interaction was assessed by examining the overlap between significant analytes. Sex-diagnosis interactions were re-analysed for overlapping analytes using age-diagnosis interaction as an additional covariate. Interactions between sex and positive, negative, and general PANSS scores for $\log_2$-transformed analyte levels in schizophrenia patients were also investigated using the same method as above. Average Pearson correlation coefficients between $\log_2$-transformed analyte levels and scores across cohorts were found using Fisher's z transformation (*391*). The FDR analogue of the *p*-value (*q*-value) was calculated for these analyses, described in **Chapter 2**. In all analyses, heterogeneity in regression coefficients across the cohorts was evaluated by assessing the impact of adding an interaction between cohort and the variable of interest to the linear regression. Evidence of heterogeneity was deemed to be present when *p*<0.05. Finally, the effect of six weeks of treatment with antipsychotic medication on serum analyte levels for 33 patients in Cohort 3 was evaluated using a paired *t*-test. See **Chapter 2** for more details on the analyses used. Gene ontology (GO) biological process (BP) terms for significant proteins were found using the UniProt website (*www.uniprot.org; accessed March 2015)* and European Bioinformatics Institute's QuickGO online database *(www.ebi.ac.uk/QuickGO; accessed March 2015)*.

5.2.5 Classification of schizophrenia patients and controls

Serum molecular data from the four cohorts of schizophrenia patients and control subjects was combined using ComBat (*303*) in the sva package (*304*). They were then used to build random forest classifiers to distinguish between patients and controls in males and females separately. For both classifiers, 5000 trees were constructed using the randomForest package in R (*329*). The out-of-bag (OOB) error of these classifications was assessed and the mean decrease in the Gini index, a measure of node purity, was used to assess variable importance. **Chapter 2** provides more details on random forests. The male classifier was then used to classify female patients and the female classifier was used to classify male patients in order to assess the

generalizability of the classifiers to patients and controls of the opposite sex. Receiver operating characteristic (ROC) curves were plotted and area under the ROC curve (AUC) was calculated.

5.3 Results

There were no significant differences in age, BMI, tobacco, or cannabis use between patients and controls in cohorts where this information was present (**Table 5.1**). Female hormonal status (OC use/menstrual cycle/postmenopausal), as predicted from the classifier in **Chapter 3**, also did not differ between patients and controls. Females were older than males, but did not differ in BMI, tobacco or cannabis consumption, or in general, positive, negative, or total PANSS scores. Patients treated with antipsychotic medication experienced a significant improvement in all PANSS scores (**Table 5.2**).

Multiplex immunoassay measurement of serum samples resulted in identification of 45 analytes with significantly different concentrations in schizophrenia patients compared to controls after accounting for selected covariates in the linear model (**Figure 5.1**). *Q*-values<0.10 were found for all significant analytes. However, 14 of these had significant heterogeneity in coefficients across cohorts. Thirty-three of the 44 analytes were previously identified as altered in schizophrenia compared to control patients in analyses of individual cohorts included in this study and in independent studies (*130*, *209*, *212*, *268*, *346*). Most analytes were present at higher levels in schizophrenia patients compared to controls. Seven of the 45 analytes identified showed significant interactions with sex.

In total, 16 log_2-transformed analytes had significant sex-diagnosis interactions in the linear model, with three [testosterone, interleukin (IL)-15, and ferritin] having q<0.10 (**Table 5.3**; **Figure 5.2**). After using untransformed (raw) analyte concentrations, six of these were no longer significant, including testosterone. Eight analytes showed significant heterogeneity in regression coefficients across cohorts. All seven male-specific analytes [IL-15, α1-antitrypsin (AAT), IL-8, macrophage derived chemokine (MDC), thyroxine binding globulin (TBG), intercellular adhesion molecule-1 (ICAM-1), and macrophage inflammatory protein-1α (MIP-1α)] and one quantitative male analyte [ferritin] were present at higher levels in male schizophrenia patients. With the exception of TBG, all were related to immune/inflammatory response. Female-specific analytes [vascular endothelial growth factor (VEGF), angiotensin-converting enzyme (ACE), stem cell factor (SCF), von Willebrand factor (vWF), prolactin, and SHBG] and testosterone (quantitative female analyte) were involved in various biological processes. These include reproduction, circulatory system and immune processes, blood coagulation, and neuron development and function. VEGF, ACE, SCF, and SHBG were reduced in female schizophrenia patients, while vWF, prolactin, and testosterone were increased. FAI also showed a female-specific increase in

schizophrenia patients. Plasminogen activator inhibitor 1 (PAI-1), which plays a role in regulating blood coagulation, also had a qualitative interaction and was increased in male schizophrenia patients relative to females. Interactions between age and diagnosis were also assessed for these analytes to account for potential confounding with the sex-diagnosis interaction due to the older age of female subjects. Only vWF had a significant age-diagnosis interaction ($p = 0.001$). The sex-diagnosis interaction for this analyte was no longer significant ($p = 0.10$) after including age-diagnosis interaction as an additional covariate.

Four analytes with significant sex-diagnosis interactions [testosterone, MDC, prolactin, and AAT] also showed significantly different associations with positive and/or general PANSS scores between male and female schizophrenia patients (**Table 5.4**) in the linear model. Despite higher levels being found in male schizophrenia patients only, serum levels of AAT and MDC were negatively associated with female positive and general PANSS scores and with female general PANSS scores, respectively. In addition, prolactin and testosterone levels, elevated in female schizophrenia patients, were negatively associated with positive and general symptoms in males.

Finally, after six weeks of various antipsychotic treatments the levels of five analytes with significant sex-diagnosis interactions [prolactin, MDC, ICAM-1, ACE, and SHBG] were changed (**Table 5.3**). The largest change in concentration was seen for prolactin (154% increase from baseline). MDC and ICAM-1 were increased after antipsychotic treatment and were also increased specifically in male schizophrenia patients. SHBG levels, reduced in female schizophrenia patients only, were also reduced after antipsychotic treatment. In contrast, serum ACE concentration was decreased specifically in female schizophrenia patients, but increased after treatment.

Good classification accuracy of patients and controls was achieved for both male (AUC = 0.90) and female (AUC = 0.92) random forest classifiers (**Figure 5.3**). In addition, the performance of the classifier built on male data was comparable when used to classify female subjects (AUC = 0.84) and vice versa (AUC = 0.89 for classification of male subjects using a classifier built using female molecular data). Seven of the twenty most important serum analytes for male and female classifiers were shared, including IL-10, sortilin, cortisol, α2-macroglobulin, luteinizing hormone (LH), carcinoembryonic antigen (CEA), and macrophage migration inhibitory factor (MIF).

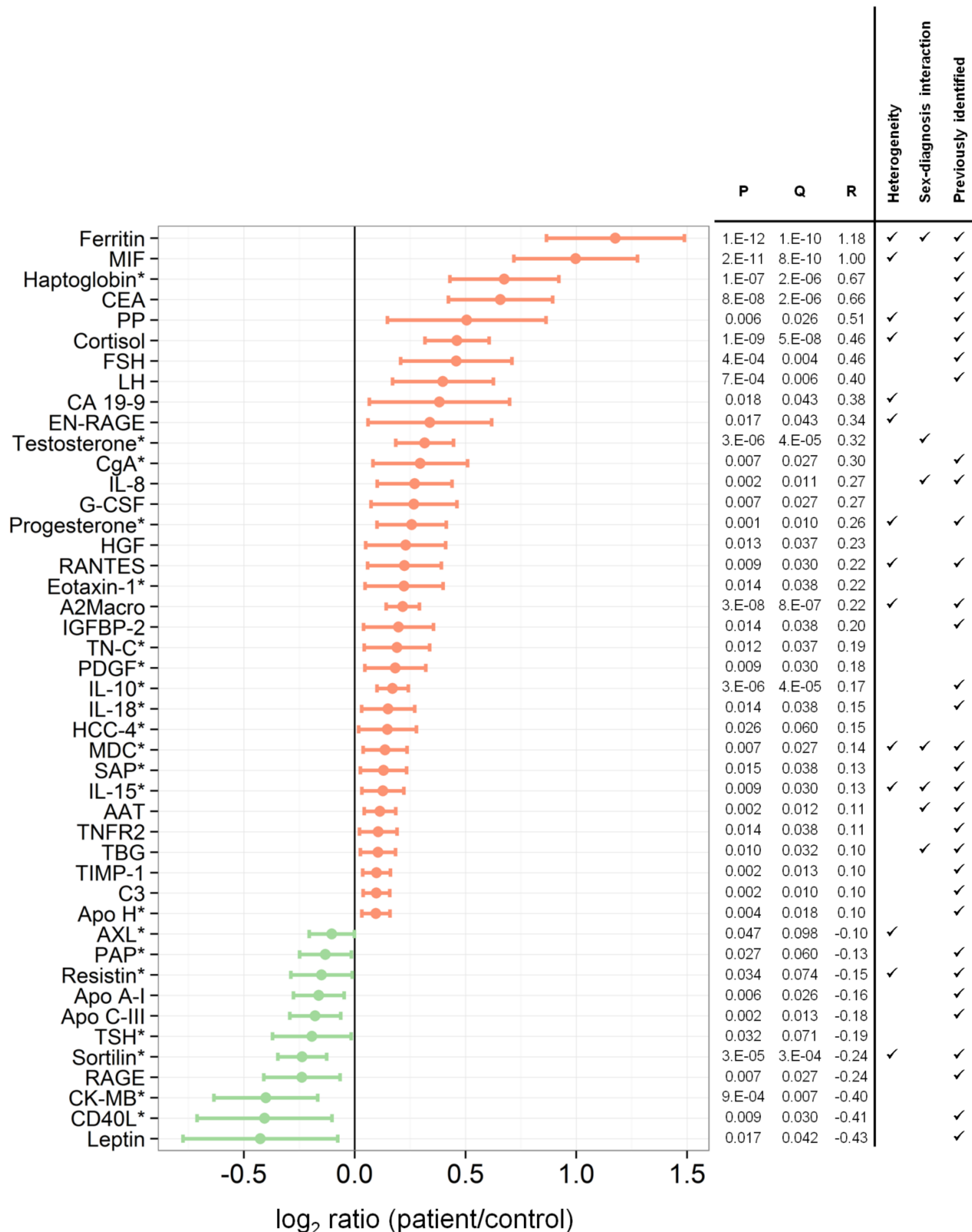
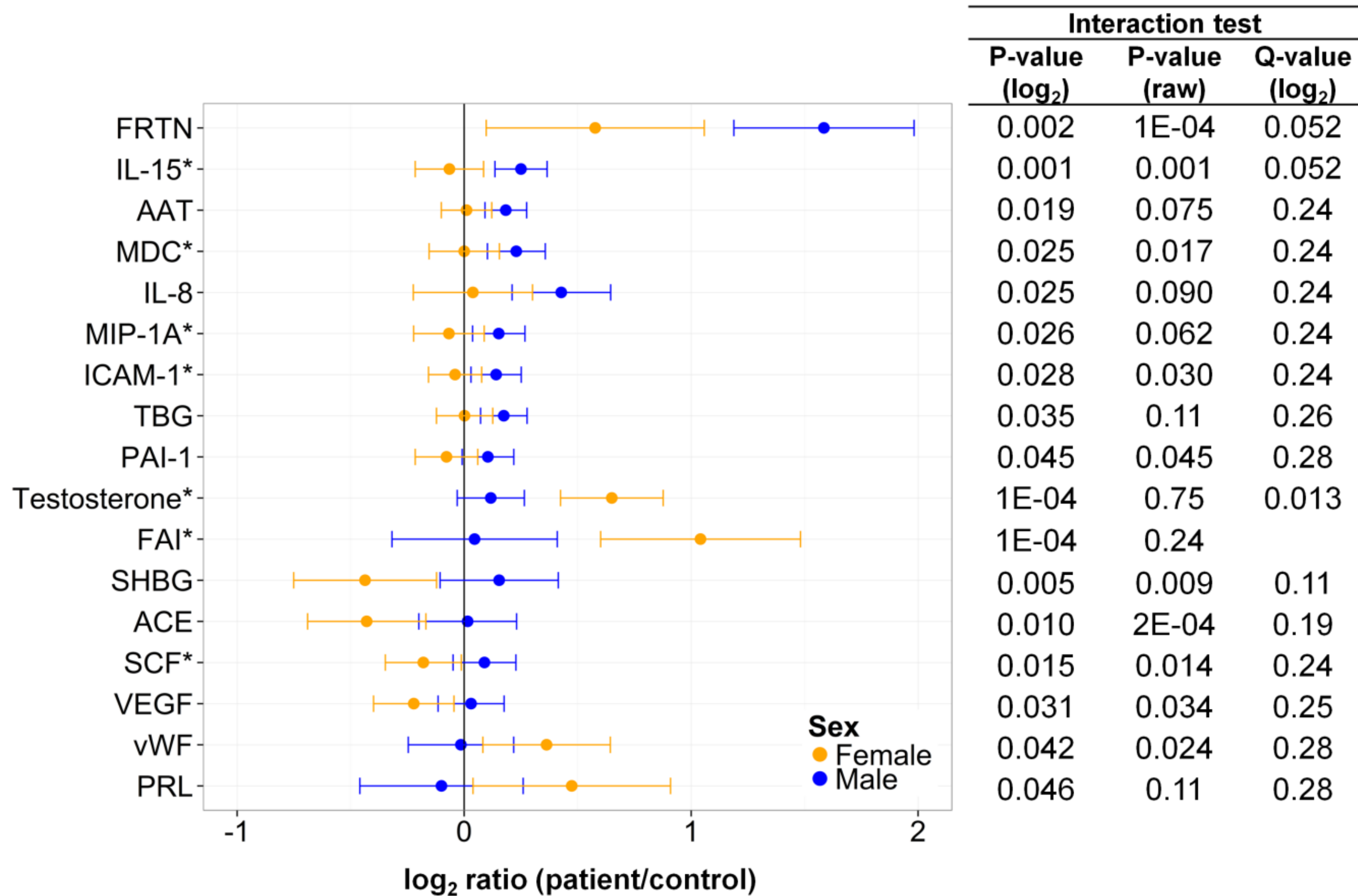


Figure 5.1. Analytes with significantly different $\log_2$-transformed serum concentration between schizophrenia patients and controls. The $\log_2$ ratios of serum molecular concentrations of schizophrenia patients compared to controls, shown in the plot with their 95% confidence intervals, were coefficients from the linear regression. Robust regression was used where * follows the analyte name. Analytes with significant sex-diagnosis interactions are marked (see **Table 5.3**/**Figure 5.2**). Analytes identified as having significantly different serum concentrations in schizophrenia in previous studies of individual cohorts included in the study or in independent cohorts are marked. Presence of cohort heterogeneity is marked. **Abbreviations:** P (*p*-value); Q (*q*-value); R ($\log_2$ ratio (patient/control)). Analyte abbreviations can be found in the **Appendix** (**Table A.1**).

Table 5.3. Log_2-transformed analytes with significant interactions between schizophrenia diagnosis and sex. Interaction *p*-values are also shown for the untransformed (raw) analyte concentrations. The log_2 ratios of serum molecular concentrations (R; patient/control) were coefficients from the linear regression. The percentage impact (%) of antipsychotic treatment on serum molecular concentrations was calculated from the regression coefficient in the linear model. Use of robust regression and presence of cohort heterogeneity for log_2 transformed analyses is marked. Orange boxes = male-specific analytes; white boxes = analytes with a qualitative interaction; blue boxes = female-specific analytes. Free androgen index (FAI) was the ratio of testosterone to SHBG. Darker colours represent a quantitative interaction for which the difference in analyte concentration between patients and controls is greater in males (dark orange) or females (dark blue). Analyte abbreviations can be found in the **Appendix** (**Table A.1**).

Analytes	Interaction		Interaction *p*-value (raw)	Females			Males			Robust regression	Heterogeneity	Treatment Main Effect	
	P-value	Q-value		R	P-value	Q-value	R	P-value	Q-value			%	P-value
Ferritin	0.002	0.052	1E-04	0.58	0.019	0.043	1.59	1.E-13	3.E-12		✓	-13.9	0.082
IL-15	0.001	0.052	0.001	-0.06	0.40	0.55	0.25	2.E-05	3.E-04	✓	✓	-8.6	0.59
AAT	0.019	0.24	0.075	0.01	0.85	0.97	0.18	1.E-04	8.E-04			-3.5	0.26
MDC	0.025	0.24	0.017	0.00	1.00	1.00	0.23	5.E-04	0.002	✓	✓	**18.5**	**2E-06**
IL-8	0.025	0.24	0.090	0.04	0.77	0.91	0.43	1.E-04	8.E-04			-13.2	0.36
MIP-1α	0.026	0.24	0.062	-0.07	0.40	0.55	0.15	0.010	0.031	✓	✓	-3.0	0.28
ICAM-1	0.028	0.24	0.030	-0.04	0.50	0.67	0.14	0.013	0.034	✓		**13.8**	**0.006**
TBG	0.035	0.26	0.11	0.00	0.97	1.00	0.17	9.E-04	0.004			4.4	0.40
PAI-1	0.045	0.28	0.045	-0.08	0.27	0.41	0.10	0.071	0.13			3.3	0.35
Testosterone	1E-04	0.013	0.75	0.65	4E-08	6.E-07	0.12	0.12	0.21	✓		2.1	0.79
FAI	1E-04		0.24	1.04	5E-06		0.05	0.80		✓			
SHBG	0.005	0.11	0.009	-0.44	0.007	0.024	0.15	0.24	0.39			**-19.2**	**0.008**
ACE	0.010	0.19	2E-04	-0.43	0.001	0.005	0.02	0.89	0.97		✓	**21.5**	**6E-05**
SCF	0.015	0.24	0.014	-0.18	0.035	0.071	0.09	0.20	0.34	✓	✓	6.2	0.10
VEGF	0.031	0.25	0.034	-0.22	0.014	0.035	0.03	0.68	0.84			-0.3	0.93
vWF	0.042	0.28	0.024	0.36	0.011	0.033	-0.01	0.91	0.97		✓	-4.4	0.63
Prolactin	0.046	0.28	0.11	0.47	0.033	0.070	-0.10	0.59	0.75		✓	**153.6**	**6E-6**



Analyte	Interaction test P-value (log_2)	Interaction test P-value (raw)	Interaction test Q-value (log_2)
FRTN	0.002	1E-04	0.052
IL-15*	0.001	0.001	0.052
AAT	0.019	0.075	0.24
MDC*	0.025	0.017	0.24
IL-8	0.025	0.090	0.24
MIP-1A*	0.026	0.062	0.24
ICAM-1*	0.028	0.030	0.24
TBG	0.035	0.11	0.26
PAI-1	0.045	0.045	0.28
Testosterone*	1E-04	0.75	0.013
FAI*	1E-04	0.24	
SHBG	0.005	0.009	0.11
ACE	0.010	2E-04	0.19
SCF*	0.015	0.014	0.24
VEGF	0.031	0.034	0.25
vWF	0.042	0.024	0.28
PRL	0.046	0.11	0.28

Figure 5.2. Log_2-transformed analytes with significant interactions between schizophrenia diagnosis and sex. Interaction *p*-values are also shown for the untransformed (raw) analyte concentrations. The log_2 ratios of serum molecular concentrations of schizophrenia patients compared to controls, shown in the plot with their 95% confidence intervals for males and females separately, were coefficients from the linear regression. Robust regression was used where * follows the analyte name. Analyte abbreviations can be found in the **Appendix** (**Table A.1**).

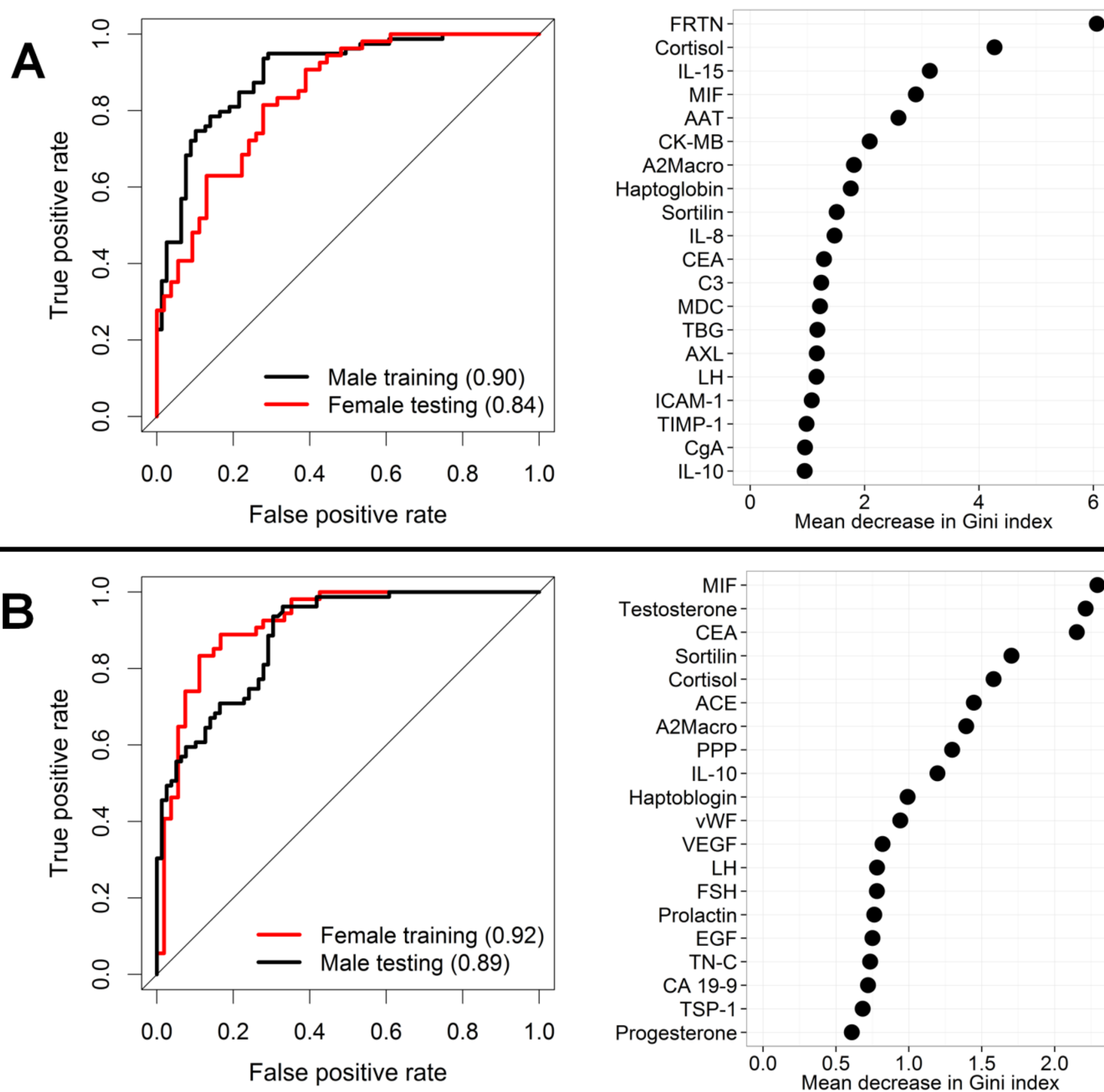


Figure 5.3. ROC curve illustrating classification of schizophrenia for (A) males and (B) females using random forests. The classifier for male schizophrenia was used to classify female schizophrenia (illustrated in **A**) and vice versa (illustrated in **B**). The legends show AUCs in brackets. The twenty most important variables for **(A)** male and **(B)** female classification are shown in the right panel. Analyte abbreviations can be found in the **Appendix** (**Table A.1**).

Table 5.4. Sex differences in associations of serum molecular concentrations with positive, negative, and general PANSS score ratings. Only analytes with significant interactions between schizophrenia diagnosis and sex were investigated (see **Table 5.3**/**Figure 5.2**). Sex-specific associations with symptoms (p<0.05 in one sex and p>0.20 in the other) are recorded for the sex with the significant association, while qualitative interactions are recorded for both sexes. Analytes without a significant interaction between symptom severity and sex are not included. The coefficient from the linear model represents the percent change in analyte concentration for a small change in symptom severity in the linear model and significant coefficients (p<0.05) are highlighted. Presence of cohort heterogeneity is included, with P_{HET} < 0.05 indicating significant cohort heterogeneity. Pearson correlation coefficients were averaged across cohorts using Fisher's z transformation. Robust regressions were used for MDC. **Abbreviations:** P (p-value); Q (q-value); HET (cohort heterogeneity); r (averaged Pearson correlation coefficient). Analyte abbreviations can be found in the **Appendix** (**Table A.1**).

	Testosterone	MDC	Prolactin	AAT
Positive symptoms	**P = 0.003, Q = 0.052, P_{HET} = 0.74**		**P = 0.016, Q = 0.088, P_{HET} = 0.060**	**P = 0.010, Q = 0.077, P_{HET} = 0.37**
Male	-1.75 (P = 0.017, Q = 0.034, r = -0.38)		-4.22 (P = 0.007, Q = 0.020, r = -0.41)	
Female	1.34 (P = 0.070, Q = 0.10, r = 0.21)			-1.18 (P = 0.002,Q = 0.015, r = -0.35)
Negative symptoms				
Male				
Female				
General symptoms	**P = 0.010, Q = 0.11, P_{HET} = 0.73**	**P = 0.014, Q = 0.11, P_{HET} = 0.039**	**P = 0.036, Q = 0.16, P_{HET} = 0.038**	**P = 0.039, Q = 0.16, P_{HET} = 0.12**
Male	-0.95 (P = 0.022, Q = 0.10 r = -0.38)		-1.83 (P = 0.041,Q = 0.11, r = -0.35)	0.17 (P = 0.42, Q = 0.42, r = 0.095)
Female	0.67 (P = 0.15, Q = 0.24, r = 0.19)	-0.74 (P = 0.025, Q = 0.10, r = -0.32)		-0.46 (P = 0.055, Q = 0.11, r = -0.17)

5.4 Discussion

This chapter investigated overall and sex-dependent differences in serum molecular concentrations in first onset, antipsychotic naive schizophrenia patients compared to controls. A number of differences were found between paranoid subtype schizophrenia patients and controls in the concentrations of hormones, growth factors, and analytes involved in immune/inflammatory and other pathways. The majority of these analytes have been identified in previous analyses of individual cohorts included in this study or in independent investigations (*130*, *209*, *212*, *268*, *346*). However, the finding of increased levels of progesterone, LH, and FSH in first onset schizophrenia contradicts *Bergemann et al (2005)* (*267*), who found low levels of progesterone and LH in chronic female schizophrenia patients. This may be due in part to the chronic course of illness and hyperprolactinemia caused by treatment with typical antipsychotic medication in most patients in the study. Together with increased levels of cortisol in first episode schizophrenia these findings suggest abnormalities in the hypothalamic-pituitary-adrenal (HPA) and hypothalamic-pituitary-gonadal (HPG) axes.

To this end, the most novel outcome of the study was the finding that 16 log_2-transformed molecules showed significant sex-diagnosis interactions. Among the most significant were findings that testosterone levels were increased and SHBG levels were decreased in first onset female schizophrenia patients. Together this meant that FAI, measuring the approximate amount of biologically available, unbound testosterone in serum, was also significantly increased specifically in females with schizophrenia. For males, higher testosterone levels were associated with lower positive and general symptom scores. Polycystic ovary syndrome (PCOS), diagnosed in females with infrequent ovulation and hyperandrogenemia, is often characterized by insulin resistance, and similar to the current study elevated LH and increased HPA axis activity (*392–395*). Studies in female sheep and Rhesus monkeys have found that fetal androgen exposure can cause adult ovarian hyperandrogenism, increased LH secretion with abnormal pulsatility, and metabolic disturbances (*396*). A genetic predisposition to excess androgen production may influence fetal, infant, and adolescent development and play an important part in PCOS's aetiology (*397*). This aetiology would be consistent with the disrupted sexual dimorphisms reported in cortical brain volumes in schizophrenia (*249*), reversed normal sexual dimorphisms in cerebral activation during processing of aversive stimuli (*251*) and during mental rotation (*252*), and other evidence of a reversed or disturbed cerebral sexual dimorphism (*250*, *277*). Together with increased prolactin levels, the current study points to more endocrine abnormalities in female schizophrenia. Prolactin secretion is inhibited by dopamine, which has been implicated in schizophrenia and in the action of antipsychotic medication.

Furthermore, there was an association in males only between lower positive PANSS scores and higher prolactin levels. This may suggest sex differences in dopamine involvement and regulation in schizophrenia. These endocrine abnormalities appear to be further influenced by six weeks of antipsychotic treatment, after which levels of prolactin were increased and levels of SHBG were decreased. Elevated prolactin levels caused by first-generation antipsychotics and certain second-generation antipsychotics are thought to contribute to menstrual disturbance and sexual dysfunction in schizophrenia patients (*265*, *398*).

Other female-specific differences included lower serum levels of ACE and VEGF. Decreased serum VEGF has previously been found in patients with childhood trauma in first episode psychosis (*399*) and both decreased and increased VEGF mRNA expression has been reported in different studies of the prefrontal cortex in schizophrenia patients (*291*, *400*). VEGF mediates angiogenesis and neurogenesis and has been shown to protect dopaminergic neurons from death in a rat model of Parkinson's disease (*401*). However, the cause of reduced serum VEGF in female schizophrenia patients in this study is unclear, as a study of brain tumors found tumor extract and serum VEGF concentrations to be independent (*402*) and a study of a genetic rat model of depression found reduced brain, but not serum VEGF levels (*403*). However, others have found a correlation between VEGF serum and cerebrospinal fluid (CSF) concentrations in neurotuberculosis patients (*404*) and have found that peripheral VEGF is required for exercise-induced adult hippocampal neurogenesis in mice (*405*). Serum ACE was also decreased in female schizophrenia patients in this study. Peripheral administration of ACE inhibitors has been shown to have effects on dopaminergic neurotransmission, as it can cross the blood-brain barrier and result in increased levels of striatal dopamine dialysate in rats (*406*), inhibition of apomorphine-induced oral stereotypy in rats similar to haloperidol (*407*), reduced effects of apomorphine on prepulse inhibition in mice (*408*), and improved dyskinesia and motor fluctuations in Parkinson's disease (*409*). The current study also showed that serum ACE levels were significantly increased following six weeks of antipsychotic treatment. However, positive correlations have been found between patient ACE levels in CSF, but not serum, and length of schizophrenic psychosis, duration of current psychotic episode, and duration of current hospitalization (*410*). Future studies should investigate how peripheral and central VEGF and ACE are correlated in schizophrenia and whether they are involved in sex-specific dopaminergic regulation in schizophrenia.

Finally, male-specific analytes were increased in schizophrenia and largely related to immune response and inflammation. This raises the possibility that some aspects of the widely-reported immunological abnormalities in schizophrenia may be specific for males. In addition to abnormalities in cytokine levels in schizophrenia (*211*), previous studies have

shown that immune-related genes, maternal infection, a higher prevalence of autoimmune disease, activated microglial cells, and increased blood-brain-barrier permeability are associated with schizophrenia (*411*). IL-15, one of the most significant findings and increased specifically in male schizophrenia patients, is expressed by macrophages and other cells, recruits and promotes the proliferation and survival of activated T lymphocytes (*412*), and induces expression of chemokines and their receptors in T lymphocytes, including MIP-1α and IL-8 (*413*) that were also increased in males with schizophrenia. This is in line with the macrophage-T-lymphocyte theory of schizophrenia proposed by *Smith and Maes (1995)* (*224*), which postulates that chronically activated macrophages and T-lymphocytes are key mediators of schizophrenia. In fact, males are more prone to autoimmune diseases characterized by acute inflammation and proinflammatory immune response involving macrophages, neutrophils, and T cells (*380*). MDC was also increased specifically in male schizophrenia patients, and is associated with chronic inflammation and produced by microglia and CNS-infiltrating leukocytes in mice with experimental autoimmune encephalomyelitis (*414*). MDC and ICAM-1 were increased after six weeks of antipsychotic treatment. Both pro- and anti-inflammatory effects have been observed for antipsychotic medication (*415*) and are thought to be associated with metabolic side effects and an increased risk of cardiovascular disease (*416*, *417*) and improvement of psychotic symptoms (*418*, *419*), respectively. Sex-specific associations between PANSS scores and levels of MDC and AAT were found, with higher levels associated with lower scores only in females. Increases in levels of AAT in females may protect against damage from inflammation and ameliorate symptoms, but further work is needed to elucidate the role of these analytes in male and female schizophrenia.

The limitations of this study should be taken into account when assessing these findings. A number of findings were inconsistent between cohorts and further work is needed to establish the causes of this heterogeneity. Fewer female patient samples were available, reducing the power to detect molecular differences in this group. In addition, potential confounding factors BMI, tobacco, and cannabis consumption were not recorded in cohort 2. Information on other prescription drug use and other demographic, lifestyle, and health variables, such as socioeconomic status, fasting status, blood pressure, and family history were unavailable. Menopausal status, OC use, menstrual cycle phase, and other variables were not recorded. Different types of medication were assessed together in the investigation of effects of antipsychotic treatment. Larger cohorts with better characterized samples will be needed for replication of these results, to assess effects of hormonal status in females, and to investigate the broader diagnosis of schizophrenia, as schizophrenia subtypes have been eliminated from the DSM-5 (*1*). Follow-up studies would account for possible misdiagnosis

and change in diagnosis over time, thought to occur more frequently in females (*238*). In addition, future studies should include greater numbers of late onset patients to facilitate studies of the effects of age on sex-dependent molecular signatures of schizophrenia as levels of estrogen in females decline. Larger follow-up studies of antipsychotic treatment would allow sex-dependent molecular changes to be assessed in different types of medication and in pre- and postmenopausal hormonal states.

The findings from this study indicate differences in the underlying molecular phenotypes of male and female paranoid subtype schizophrenia patients. They include elevated levels of certain inflammatory markers specifically in male schizophrenia and endocrine abnormalities in female schizophrenia patients. With the exception of vWF, these were not due to differences in age at first episode between males and females. This would have been predicted by the hypothesis that males are more prone to a neurodevelopmental, early onset schizophrenia, while females are more prone to a late onset schizophrenia similar to affective psychosis (*273–276*). Some of these analytes also showed sex differences in their associations with symptomatology and were modified by antipsychotic treatment. However, consistent with the high degree of overlap in presentation of male and female schizophrenia, the majority of analyte concentrations different in schizophrenia were not significantly modified by sex. Future studies will be needed to validate these results and further research should be conducted on male- and female-specific aetiologies, pathophysiologies, and adaptive mechanisms in schizophrenia in order to elucidate the role of these analytes. Furthermore, sex-specific treatment options, in line with these findings, should potentially be considered. A recent meta-analysis showed estrogen has a beneficial effect in female schizophrenia (*9*). Metformin is a treatment for PCOS and has been found to restore menstruation and reduce levels of prolactin, LH, and testosterone in women with antipsychotic-induced amenorrhea during first-episode schizophrenia (*420*). Further studies should assess how metformin treatment may affect symptoms of schizophrenia. Non-steroidal anti-inflammatory drug (NSAID) (*421*) and glucocorticoid use (*422*) has been associated with a lower risk of psychosis in males only. A meta-analysis of anti-inflammatory efficacy in schizophrenia showed significant improvement in symptoms with aspirin and *N*-acetylcysteine, but these were not stratified by sex (*423*).

Chapter 6 A meta-analysis of gene expression profiling of the prefrontal cortex: schizophrenia, sex, and sex-schizophrenia diagnosis interaction effects

6.1 Introduction

Post-mortem microarray studies in schizophrenia have focused on the prefrontal cortex (PFC), temporal cortex, cingulate cortex, hippocampus, thalamus, and cerebellum as brain regions that may be involved in its pathophysiology. The PFC in particular has been the subject of a number of microarray studies that have revealed disturbed processes in schizophrenia related to synaptic function, neuronal development, mitochondrial function, immune and inflammatory functions, oligodendrocyte and astrocyte function, growth factor signalling, myelination, metal ion binding, signal transduction, sphingolipid metabolism, and others (*51*, *198*, *199*, *203*, *204*, *214*, *291*, *292*, *424–428*). Two meta-analyses of Affymetrix HG-U133A/HG-U133 Plus 2.0 microarray studies of Brodmann area (BA) 9, 10, and 46 of the PFC in schizophrenia have been conducted by *Pérez-Santiago et al (2012)* (*429*) (six studies) and *Mistry, Gillis, and Pavlidis (2012)* (*430*) (seven studies, six overlapping with by *Pérez-Santiago et al (2012)* (*429*)). These have revealed a number of differentially expressed genes in schizophrenia with q<0.1 involved in processes including neuronal cell-cell signalling, mesenchymal induction, and mitogen-activated protein kinase (MAPK) signalling [*Pérez-Santiago et al (2012)* (*429*)] and neuronal communication, processes affected by synaptic functioning, metabolic processes, ubiquitination, and immune function [*Mistry, Gillis, and Pavlidis (2012)* (*430*)]. However, due to the use of different datasets, data pre-processing, outlier removal, analysis, and/or false discovery rate (FDR) procedures between the meta-analyses, only 19 of the findings overlapped between the 160 significant probe sets in *Pérez-Santiago et al (2012)* (*429*) and 125 significant probe sets in *Mistry, Gillis, and Pavlidis (2012)* (*430*). These meta-analyses did not make use of robust regression in the presence of regression outliers, array weights to deal with arrays of variable quality, or moderated t-statistics to potentially further increase power and stability of the meta-analysis. In addition, these and other microarray studies have failed to examine potential sex and sex-diagnosis interaction effects on gene expression.

Sex differences in the neuroanatomy and function of the PFC are profound. Cognitive processes such as verbal fluency and certain types of visuospatial skills also differ between males and females (*431*). Neurotransmitter systems differ between males and females and are further discussed in **Chapter 1**. Normal adult PFC volumes are sexually dimorphic, in line with areas having high levels of estrogen and androgen receptors during brain development in animal studies (*432*). Functional magnetic resonance imaging (fMRI) studies have also shown more bilateral and right hemisphere activation in male subjects and left hemisphere activation in female subjects (*433*), and greater intensity changes in females in the middle, inferior, and orbital PFC during working memory tasks (*434*). Furthermore, estrogen-dependent stress-induced PFC dysfunction occurs in rats, as measured by a working memory task (*435*). Previous microarray studies of sex-biased gene expression in the PFC have revealed sex differences in expression of genes located on the sex chromosomes (*436–438*) and in expression of autosomal genes involved in transcription, intracellular signalling, and heat shock proteins (*437*).

As described in **Chapter 1,** schizophrenia is characterized by marked sex and gender differences in clinical presentation including a higher male incidence of schizophrenia, older female age of onset, differences in symptomatology, better response to antipsychotic medication in premenopausal females, and better course of illness in women compared to men (*238*). **Chapter 5** showed that some serum molecular concentrations varied differently with schizophrenia diagnosis in male and female antipsychotic naive, first onset patients of the paranoid subtype. In female patients, total and free serum testosterone, prolactin, and von Willebrand factor were elevated, while ACE and VEGF were reduced. On the other hand, a number of serum analytes involved in immune and inflammatory response were elevated only in male patients. Evidence for sex-specific and sex-dependent genetic risks has emerged (*271*), including genes involved in dopaminergic metabolism (*272*). Sex differences in neuroanatomical and functional brain abnormalities in the PFC have been reported in schizophrenia (see **Chapter 1**), including smaller dorsomedial grey matter (*255*) and frontomedial cortex volumes (*249*) in male schizophrenia compared to male controls and smaller female orbital cortex (*255*) and dorsolateral prefrontal cortex grey matter (*256*) volume in schizophrenia relative to opposite sex comparisons. Other findings include disturbed sex dimorphisms in brain activations during mental rotation and processing of aversive stimuli in areas of the PFC (*251*, *252*) and sex differences in the relationship between frontal lobe volume and disorganization and suspicion-hostility symptoms in schizophrenia (*257*). Cognitive performances of male and female schizophrenia patients may also differ, but there is conflicting evidence for this (*439*, *440*). Taken together, these findings suggest sex-specific pathophysiological mechanisms may occur in the PFC in

schizophrenia. However, to date no microarray study has investigated this possibility, potentially due to the small number of female subjects in individual cohorts.

6.1.1 Aims and objectives

The investigations presented in this chapter had the following aims and objectives:

1.) To assess potential pathophysiological mechanisms in schizophrenia by evaluating overall and sex-dependent differences in gene expression in the PFC in schizophrenia patients compared to controls. Robust regression, array weights, and moderated *t*-statistics were used in order to improve upon previous analyses.

2.) To examine the effects of possible confounding factors on gene expression, including lifetime antipsychotic medication use, suicide status, alcohol consumption, and drug use.

3.) To investigate potential sex differences in the function of the PFC and their implications in mental disorders by assessing sex-biased gene expression in control subjects.

4.) To assess overrepresented gene ontology (GO) biological process (BP) terms in PFC findings in order to gain insight into affected processes in each analysis of differential gene expression.

5.) Finally, to compare PFC findings to serum results obtained in **Chapters 3** and **5** in order to examine the link between the periphery and brain**.**

A meta-analysis of five microarray studies containing 238 subjects in post-mortem PFC tissue (BA 9, 10, 46) was performed in order to achieve these aims. By combining several microarray studies, it was hoped that sufficient power could be obtained to detect significant sex-dependent differences in gene expression in schizophrenia. This was particularly important given the small number of females included in these microarray studies.

6.2 Materials and Methods

6.2.1 Clinical samples

Five microarray datasets of schizophrenia and control prefrontal cortex (PFC; BA 9, 10, 46) were downloaded from the Stanley Medical Research Institute (SMRI; two datasets *(www.stanleygenomics.org, accessed December 2014)*, Harvard Brain Tissue Resource

Centre (HBTRC – Mclean66 collection, *national_databank.mclean.harvard.edu, accessed December 2014)*, and Gene Expression Omnibus (GEO) at the National Center for Biotechnology (two datasets; *www.ncbi.nlm.nih.gov/geo, accessed December 2014)*. **Table 6.1** contains detailed demographic and additional information for each of these datasets and **Figure 6.1** shows a lateral view of these regions in the PFC. Demographic and other information was obtained from the data repository or from articles published on the individual datasets (*198*, *214*, *291–293*).

Datasets were identified from previous meta-analyses by *Mistry, Gillis, and Pavlidis (2012)* (*430*) (seven studies) and *Pérez-Santiago et al (2012)* (*429*) (six studies, overlapping with the studies used in *Mistry, Gillis, and Pavlidis (2012)* (*430*)). However, raw data for two of the studies used in the *Mistry, Gillis, and Pavlidis (2012)* (*430*) meta-analysis (named *Mirnics* and *Haroutunian* in the article) could not be obtained from the authors. Multiple microarray SMRI datasets were available, but represented repeated measurements of subjects, so here *Kato* and *Altar C* datasets were selected for the two independent SMRI brain collections (Array and Consortium collections, respectively). In contrast to the current study, *Mistry, Gillis, and Pavlidis (2012)* (*430*) chose the *Bahn* dataset and *Pérez-Santiago et al (2012)* (*429*) did not state which dataset was used to represent the SMRI Array collection. Manual searches were also conducted in GEO, ArrayExpress at the European Bioinformatics Institute (http://www.ebi.ac.uk/arrayexpress/), and the Stanford Microarray Database (http://smd.princeton.edu/), but these did not yield any new datasets. Only data collected using Affymetrix (Santa Clara, CA, USA) HG-U133A or HG-U133 Plus 2.0 chips were included.

Table 6.1. Patient and control characteristics by cohort. Values are shown as mean ± sd. **(A)** presents a summary of data available in all cohorts. P-values in the last column are from a Welch's *t*-test of differences between male/female patients and controls ($p<0.05$ considered significant). **(B)** presents a partial summary of data available in only some of the cohorts. Age is reported in years as (21-40 : 41-60 : > 61-80) in the Harvard cohort; smoking at time of death is reported as (yes : no : unknown); lifetime alcohol use is reported as (present (heavy/moderate use) : past (heavy/moderate use) : social : none); lifetime drug use is reported as (present (heavy/moderate use) : past (heavy/moderate use) : social : little or none : unknown); ethnicity is reported as (white : non-white : unknown); and suicide status is reported as (yes : no : unknown). **Abbreviations:** PMI (post-mortem interval); M (male); F (female); BA (Brodmann area); SMRI (Stanley Medical Research Institute); HBTRC (Harvard Brain Tissue Resource Centre); P (*p*-value).

(A)	**Altar C (SMRI Consortium collection)**	**Kato (SMRI Array collection)**	**GSE17612**	**GSE21138**	**Harvard**	**Overall**	**P**
Controls N (M/F)	11 (7/4)	33 (24/9)	23 (12/11)	29 (24/5)	25 (17/8)	121 (84/37)	
Patients N (M/F)	10 (7/3)	35 (26/9)	28 (19/9)	29 (24/5)	15 (12/3)	117 (88/29)	
Controls age (M/F)	49.0±7 / 35.8±6	46.0±8 / 40.0±5	63.0±25 / 75.6±16	43.6±15 / 50.2±22	7:5:5 / 0:1:7		0.48 /
Patients age (M/F)	39.1±13 / 50.0±17	41.3±8 / 46.3±9	72.9±13 / 74.2±20	43.3±15/45.8±26	3:5:4 / 0:0:3		0.69
Controls brain pH (M/F)	6.3±0.2 / 6.2±0.3	6.7±0.2 / 6.5±0.3	6.5±0.3 / 6.5±0.3	6.3±0.2 / 6.3±0.1	6.5±0.3 / 6.3±0.3	6.5±0.3 / 6.4±0.3	5E-05 /
Patients brain pH (M/F)	6.1±0.3 / 6.1±0.2	6.4±0.3 / 6.6±0.2	6.2±0.2 / 6.0±0.2	6.2±0.2 / 6.3±0.1	6.5±0.2 / 6.2±0.4	6.3±0.3 / 6.3±0.3	0.18
Controls PMI (M/F)	18.9±9 / 32.5±10	28.3±12 / 33.9±15	11.7±4 / 8.0±4	39.1±15 / 46.8±8	22.1±4 / 18.0±7	27.0±14 / 24.3±16	0.89 /
Patients PMI (M/F)	34.1±16 / 37.3±20	30.8±17 / 33.2±11	9.4±7 / 7.2±6	39.0±12 / 43.4±12	20.8±5 / 23.0±1	27.3±17 / 26.3±17	0.64
Batches (batches used)*	8 (7)	13 (12)	3	6 (5)	8	38 (35)	
Brain region(s)	Frontal BA46/10	Frontal BA46	Anterior prefrontal cortex (BA10)	Frontal BA46	Prefrontal cortex (BA9)		
Microarray platform	HG-U133A	HG-U133A	HG-U133 Plus 2.0	HG-U133 Plus 2.0	HG-U133A		
References	*Torrey et al (2000)* (*293*) *SMRI Database*	*Iwamoto, Bundo, and Kato (2005)* (*292*) *SMRI Database*	*Maycox et al (2009)* (*291*)	*Narayan et al (2008)* (*198*)	*Glatt et al (2005)* (*214*) *Maycox et al (2009)* (*291*) *HBTRC*		

*Batches consisting of single samples were removed in order to combine data using ComBat. *Altar C, Kato,* and *Harvard* cohorts also consisted of bipolar disorder, depression, and other samples, which were removed prior to analysis. Duplicate arrays were present in *Altar C*, which were also removed prior to analysis.

(B)	Altar C (SMRI)	Kato (SMRI)	GSE17612	GSE21138	Harvard
Antipsychotics (M/F)	4000-100,000 (37,400±40,000) / 6000-50,000 (30,300±22,000) (p=0.73) [Lifetime antipsychotic use in fluphenazine mg equivalents]	50-350,000 (90,200±92,000) / 600-400,000 (70,100±130,000) (p=0.67) [Lifetime antipsychotic use in fluphenazine mg equivalents]	Doses relatively low; only four patients took doses >750mg/day chlorpromazine equivalents. Patients were elderly and most were treated long term after first admission; controls were from community/ hospital/ nursing home	50-750 (330±200) / 100-750 (390±300) (p=0.73) [chlorpromazine equivalents; incorporate lifetime exposure]	Not available.
Controls Smoking (M/F)	2:2:3 / 1:0:3	6:7:11 / 2:2:5		Not reported; data divided into short, intermediate, and long-term illness. Suicide reported to be higher in short- and intermediate duration illness compared to long-duration illness and controls.	
Patients Smoking (M/F)	3:2:2 / 1:2:0	17:4:5 / 6:0:3			
Controls Alcohol (M/F)	2:1:3:1 / 0:0:2:2	0:2:12:10 / 2:1:0:6			
Patients Alcohol (M/F)	2:1:2:2 / 0:2:1:0	11:4:5:6 / 1:2:2:4			
Controls Drugs (M/F)	0:1:0:6:0 / 0:0:0:4:0	1:0:3:20:0 / 0:0:1:8:0			
Patients Drugs (M/F)	1:0:0:6:0 / 0:1:0:2:0	9:4:4:8:1 / 0:2:0:6:1			
Controls Ethnicity (M/F)	6:1:0 / 4:0:0	24:0:0 / 9:0:0	12:0:0 / 11:0:0		5:0:12 / 3:0:5
Patients Ethnicity (M/F)	6:1:0 / 2:1:0	25:1:0 / 9:0:0	19:0:0 / 9:0:0		10:0:2 / 3:0:0
Controls Suicide (M/F)	0:7:0 / 0:4:0	0:24:0 / 0:9:0	Causes of death similar in patients and controls (broncho-pneumonia, carcinoma, cardiovascular causes most common)		Not available.
Patients Suicide (M/F)	2:5:0 / 1:1:1	4:22:0 / 3:6:0			

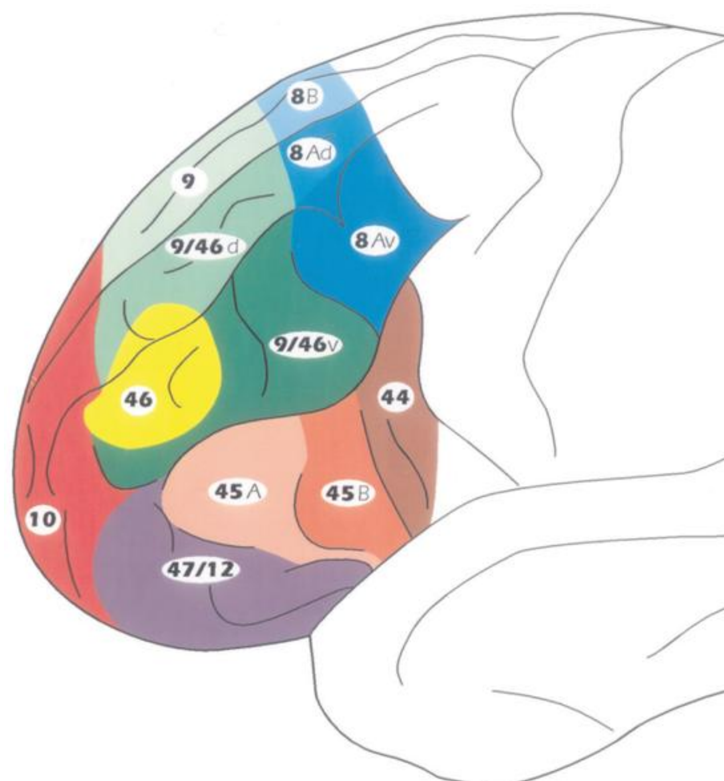


Figure 6.1. Lateral surface of the human PFC. Regions are numbered and coloured according to Brodmann area. *Image courtesy of Petrides and Pandya (2009)* (*441*)*; reprinted with permission.*

6.2.2 Data pre-processing

Data pre-processing and analysis was carried out using Bioconductor (v3.0) (*294*) with the statistical programming software R (v3.1.2) (*285*). Array quality was assessed in each dataset with the R Bioconductor arrayQualityMetrics package (*295*), using Relative Log Expression (RLE) plots, Normalized Unscaled Standard Error (NUSE) plots, and spatial distribution of feature intensities. The affy (*296*) and simpleaffy (*297*) packages were used to further examine array quality measures of ribonucleic acid (RNA) degradation, average, minimum, and maximum background levels, percentage of genes present, scale factors, 3'/5' ratios for beta-actin and glyceraldehyde 3-phosphate dehydrogenase (GAPDH), and spike-in hybridization controls. A number of poor quality arrays were found, and therefore array quality weights were calculated (*302*) after pre-processing using the limma package (*300*, *301*) (described in **Chapter 2**), and used in subsequent analyses. Raw data were $\log_2$ transformed, background corrected, normalized, and summarized using frozen robust multiarray analysis (fRMA) as described in **Chapter 2** using the frma package (*298*). Batch information was found from the scan date in the CEL files and arrays scanned on different days were regarded as different batches. Considerable variability in expression was found between batches and between different datasets, which was removed using ComBat (*303*)

in the sva package (*304*). For this final step, only probe sets common to the HG-U133A and HG-U133 Plus 2.0 chips were used. Where duplicate arrays were processed, the highest quality array was chosen from the calculated array quality weights. Bipolar disorder, depression, and other samples were removed from the SMRI and Harvard datasets to leave only schizophrenia and control subjects. In total, expression values for 22,277 probe sets in 238 subjects were analyzed. Gene symbols, gene names, and chromosome/band were mapped from Affymetrix HG-U133A probe set identifiers using the R Bioconductor package biomaRt (*336*, *337*). Gene names and symbols for any remaining unidentified probe sets were found by manual search using the Affymetrix NetAffx[TM] Analysis Centre (*442*) *(www.affymetrix.com/analysis/netaffx, accessed February 2015)*.

6.2.3 Data analysis

The limma package was used to identify significant main effects of schizophrenia diagnosis and sex and their interaction in probe set intensities (*301*, *443*). Control subjects only were used to identify significant sex-biased gene expression. Brain pH, post-mortem interval (PMI) measured in hours, sex, and decade of age at death were used as additional covariates. Robust regression (MM-estimator) was used where regression outliers were present (with Bonferroni-corrected p-value < 0.05), calculated using the robustbase package in R (*317*). P-values were calculated from moderated t-statistics (further details in **Chapter 2**) and q-values were found as described in **Chapter 2**. The criterion for significance used was q-value < 0.10. Analytes with significant sex-diagnosis interaction terms were tested separately in males and females for each disorder and classified as either male-specific, female-specific or qualitative markers. Male-specific markers were significant in males only (p<0.05 in males; p>0.20 in females) and female-specific markers were significant in females only (p<0.05 in females; p>0.20 in males). Qualitative markers had opposite associations with diagnosis between males and females. In all analyses, heterogeneity in regression coefficients across studies was evaluated by assessing the impact of adding an interaction between cohort and the variable of interest to the linear regression. Evidence of heterogeneity was deemed to be present when p<0.05.

The overlap between probe sets with a significant main effect of schizophrenia diagnosis in the present study and those reported in the individual cohorts and in the *Mistry, Gillis, and Pavlidis (2012)* (*430*) and *Pérez-Santiago et al (2012)* (*429*) meta-analyses was evaluated. Lists of significant probe sets from each cohort were found in the publications by *Maycox et al (2009)* (*291*) for the Harvard and GSE17612 cohorts; *Glatt et al (2005)* (*214*) for the Harvard cohort; *Narayan et al (2008)* (*198*) for the GSE21138 cohort; and from the lists of

'regulated genes' on the SMRI website (*www.stanleygenomics.org, accessed December 2014).*

6.2.4 Potential confounders

Potential confounding variables, including lifetime use of antipsychotic medication, suicide status, lifetime alcohol consumption, and lifetime drug use, were not reported consistently across cohorts. Therefore, the limma package (*301*, *443*) was used to identify effects of these variables on gene expression using the SMRI datasets (*Altar C* and *Kato),* where this information could be obtained. Lifetime antipsychotic use was measured in fluphenazine milligram equivalents and its effect on gene expression was evaluated in schizophrenia and bipolar patients (N=88). Suicide status was assessed as 'yes' or 'no', and lifetime alcohol/drug use was assessed as 'moderate or severe' (past/present) or 'little or none' and their effects on gene expression were measured in schizophrenia, depression, bipolar disorder, and control subjects (N=143). Potential confounding was considered when $p<0.05$.

6.2.5 Comparison with serum results

In **Chapter 5**, main effects of schizophrenia diagnosis and sex-diagnosis interactions were evaluated for serum concentrations of analytes in first onset, antipsychotic naive schizophrenia patients compared to controls. Variations in serum analyte concentrations with sex and female hormonal status were determined in **Chapter 3**. The overlap between these analytes and the present microarray findings were assessed. For this, serum analyte overlap was considered for those proteins without a significant effect of female hormonal status or where the direction of effect of hormonal status was consistent compared to males. Where multiple probe sets representing a gene were present, the most significant finding was reported together with its Affymetrix identifier.

6.2.6 GO terms

GO terms for significant genes were found using the R Bioconductor package biomaRt (*336*, *337*) and these were mapped to all ancestor terms using the GOSim package (*338*). A summary of GO terms for each significant gene was prepared, with GO terms selected based on relevance and number of annotated genes. Overrepresented GO BP terms for lists of probe sets with significant effects of sex, diagnosis, and sex-diagnosis interactions were assessed using the GOstats R Bioconductor package (*339*). A hypergeometric test that conditions on significant child terms was used, requiring at least ten annotated genes for the term and $p<0.001$.

Further information on microarray technology and analysis can be found in **Chapter 2.**

6.3 Results

The majority of patient and control samples were male (75% of patients, 69% of controls, **Table 6.1**), varying among cohorts from 61% male (GSE17612) to 83% male (GSE21138) samples. There were no significant differences in sex, age, or PMI between schizophrenia patients and controls. Male brain pH was lower in schizophrenia patients compared to controls. Most subjects across cohorts were white (66%), with the remainder being non-white (2%) or unknown/unrecorded (32%). Other variables were not reported consistently across cohorts, but in the SMRI *Altar C* and *Kato* cohorts, lifetime antipsychotic use, smoking at time of death, lifetime alcohol and drug use, and suicide status were available. Smoking at time of death was unknown for a substantial proportion of subjects (36%), but was more prevalent among patients when it was recorded (77% and 50% of recorded patients and controls, respectively). Lifetime alcohol and drug use was known for all but two SMRI subjects. Present or past moderate/severe alcohol and drug use was more prevalent in patients (51% alcohol users and 40% drug users) compared to controls (18% alcohol users and 5% drug users). Lifetime antipsychotic use was not different between male and female patients in the cohorts where this was reported.

A total of 38 batches were identified (from scan dates in CEL files) within the microarray datasets. In order to combine the batches within each of the five cohorts using ComBat, three batches consisting of only one sample were removed. A summary of array quality weights calculated and used in subsequent analyses is shown for all cohorts in **Figure 6.2.** Manual verification confirmed that low quality arrays identified from the procedure carried out in **Section 6.2.2** received low weights. Approximately 4-5% of probes were investigated using robust regression.

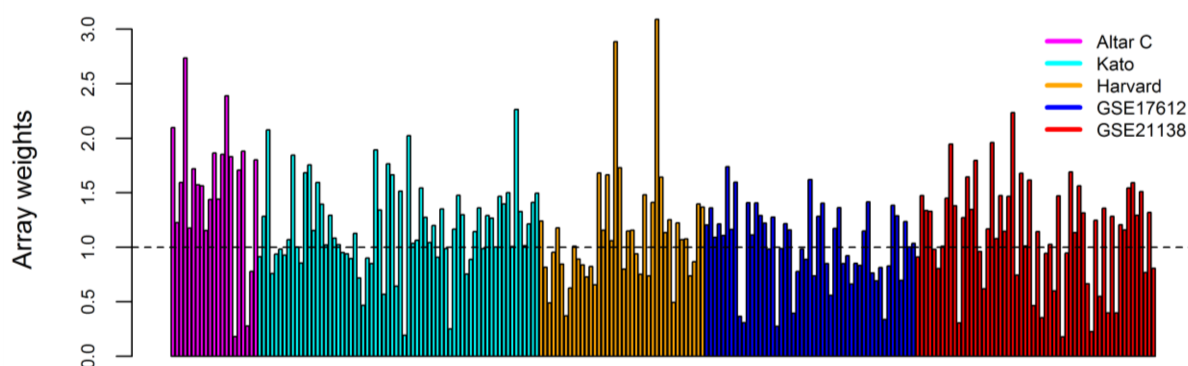


Figure 6.2. **Array quality weights used for data analysis.** Array quality weights were calculated using the limma Bioconductor package in R.

6.3.1 Differential PFC gene expression in schizophrenia

A total of 209 unique genes [229 probe sets (86 upregulated; 143 downregulated)] were significantly differentially expressed in schizophrenia compared to control PFC (q<0.1) in the linear model described in **Section 6.2.3**, as shown in **Table 6.2** and **Figure 6.3**. Repeating the analysis without using array weights and robust regression identified only 17 of these significant probe sets, as shown in **Table 6.3.** *P*-values found without using array weights are also shown in **Table 6.2**. Evidence for heterogeneity between studies was found in 32 (14%) of these findings.

Thirty-seven of the significant probe sets from *Mistry, Gillis, and Pavlidis (2012)* (*430*) and 44 probe sets from *Pérez-Santiago et al (2012)* (*429*) overlapped with findings in this study (*p*-values of overlap were less than 2E-16). In total, 68 (30%) of the current findings overlapped with significant probe sets from at least one of the meta-analyses and 13 (12 genes) agreed across all studies. The genes common to all studies were TNFSF10, FBXO9, EGR1, CA4, LPL, P4HA1, ABCA1, CYP26B1, TMEM176A, TMEM176B, ETV5, and MPPED2. Possible causes for discrepancies between the meta-analyses are further discussed in **Section 6.4.1.** Furthermore, 103 (45%) of the significant probe sets overlapped with the findings listed in one or more of the individual cohort publications or on the SMRI website (**Table 6.2**). Most consistent with the current meta-analysis was the GSE17612 cohort (46 overlapping probes) and least consistent was the GSE21138 cohort (2 overlapping probes). This lack of consistency was likely because only the top 50 differentially expressed genes were available from *Narayan et al (2008)* (*198*) and because findings were divided by short, intermediate, and long duration of illness.

Significant genes were associated with over 3000 unique GO BP terms, including metabolic processes (GO:0008152); developmental processes (GO:0032502); cell communication and signal transduction (GO:0007154 and GO:0007165); transport (GO:0006810); growth and cell proliferation (GO:0040007 and GO:0008283); response to stress (GO:0006950); cell death (GO:0008219); defence, immune, and inflammatory response (GO:0006952, GO:0006955, and GO:0006954); nervous system development (GO:0007399); synaptic transmission (GO:0007268); and others (**Table 6.2**). A GO BP term analysis revealed 12 enriched terms (**Table 6.4**). These included terms related to steroid hormone, estradiol, metal ion, lipid, alcohol, and cyclic adenosine monophosphate (cAMP) response, regulation of neuronal apoptosis, and mitochondrial function. Genes involved in these enriched functions comprised 35% of the differentially expressed genes.

Table 6.2. Probe sets with significantly different intensity between schizophrenia patients and controls in the PFC. In **(A)** probe set intensity is higher in schizophrenia, and in **(B)** probe set intensity is lower. The percentage impact of schizophrenia on serum molecular concentrations (%) from the linear regression is reported and use of robust regression is marked for each probe set. *P*-values are also reported for analyses without array weights. The direction of effect of potential confounders is marked with arrows. GO terms were used to assess participation of a gene in each biological process. The overlap of these findings with those of previous meta-analyses and individual cohorts is highlighted in blue. Cohorts are numbered (Altar C = 1; Kato = 2; GSE17612 = 3; GSE21138 = 4; Harvard = 5). **Abbreviations:** P (*p*-value); Q (*q*-value); A (lifetime alcohol use); AP (lifetime antipsychotic use); D (lifetime drug use); S (suicide); MP (metabolic process); DP (developmental process); CC & ST (cell communication and signal transduction); IR (defence/immune/inflammatory response); CP (cell proliferation); NSD (nervous system development). **GO terms (left to right):** GO:0008152; GO:0032502; GO:0007154 and GO:0007165; GO:0006810; GO:0040007 and GO:0008283; GO:0006950; GO:0008219; GO:0006952 and GO:0006955 and GO:0006954; GO:0007399; GO:0007268.

Affymetrix ID	Gene symbol	Gene name	P	Q	%	P (no array weights)	Robust regression	Heterogeneity	Potential confounders	Mistry, Gillis, and Pavlidis (2012)	Pérez-Santiago et al (2012)	Agreement with individual cohorts	MP	DP	CC & ST	Transport	Growth / CP	Response to stress	Cell death	IR	NSD	Synaptic transmission
202917_s_at	S100A8	S100 calcium binding protein A8	1.E-07	7.E-04	92.2	6.E-07	✓	✓	S↓ AP↑		✓	1,3	✓		✓	✓	✓	✓	✓	✓		
217911_s_at	BAG3	BCL2-associated athanogene 3	2.E-06	0.007	32.2	2.E-05			S↓		✓	3	✓	✓	✓				✓		✓	
204326_x_at	MT1X	metallothionein 1X	3.E-06	0.007	33.4	2.E-05					✓	2			✓		✓	✓				
202975_s_at	RHOBTB3	Rho-related BTB domain containing 3	3.E-06	0.007	22.6	7.E-05		✓		✓		2	✓		✓	✓						
200907_s_at	PALLD	palladin, cytoskeletal associated protein	3.E-06	0.007	18.8	2.E-05			S↓		✓	2,5										
212332_at	RBL2	retinoblastoma-like 2	4.E-06	0.007	9.2	4.E-05						5	✓									
208451_s_at	C4B	complement component 4B	1.E-05	0.013	24.6	9.E-04		✓	AP↑		✓	2	✓			✓		✓		✓		
216048_s_at	RHOBTB3	Rho-related BTB domain containing 3	1.E-05	0.015	18.9	5.E-04		✓		✓		2	✓		✓	✓						
59625_at	NOL3	nucleolar protein 3 (apoptosis repressor with CARD domain)	1.E-05	0.015	8.6	1.E-04			AP↑	✓			✓					✓	✓			
205967_at	HIST1H4A	histone cluster 1, H4a	3.E-05	0.025	12.2	5.E-04							✓	✓								
204060_s_at	PRKX/PRKY	Protein kinase, X/Y-linked	5.E-05	0.035	10.1	8.E-04				✓			✓	✓			✓					
217546_at	MT1M	metallothionein 1M	5.E-05	0.035	36.0	8.E-04			S↓		✓						✓					
221567_at	NOL3	nucleolar protein 3 (apoptosis repressor with CARD domain)	8.E-05	0.040	6.5	0.004							✓					✓	✓			
208581_x_at	MT1X	metallothionein 1X	9.E-05	0.041	22.7	2.E-04					✓	2			✓		✓	✓				
202481_at	DHRS3	dehydrogenase/reductase (SDR family) member 3	9.E-05	0.041	8.8	3.E-04			S↓		✓		✓	✓	✓							
214428_x_at	C4B	complement component 4B	9.E-05	0.041	17.3	0.003		✓	AP↑		✓		✓			✓		✓		✓		
212188_at	KCTD12	potassium channel tetramerization domain containing 12	9.E-05	0.041	13.2	6.E-04						3			✓							
201841_s_at	HSPB1	heat shock 27kDa protein 1	1.E-04	0.041	22.6	4.E-04	✓		S↓		✓	3	✓	✓	✓			✓	✓			
212859_x_at	MT1E	metallothionein 1E	1.E-04	0.041	16.9	2.E-04		✓	S↓		✓	2					✓					
210800_at	TIMM8A	translocase of inner mitochondrial membrane 8 homolog	1.E-04	0.041	4.6	4.E-04																

		A (yeast)																				
200897_s_at	PALLD	palladin, cytoskeletal associated protein	1.E-04	0.041	15.9	1.E-04			S↓		✓	2										
214250_at	NUMA1	nuclear mitotic apparatus protein 1	1.E-04	0.046	6.8	4.E-04																
216195_at	ANK2	ankyrin 2, neuronal	1.E-04	0.046	11.2	6.E-05						2										
216336_x_at	MT1P3	metallothionein 1 pseudogene 3	1.E-04	0.046	15.1	4.E-04			S↓		✓	2										
203318_s_at	ZNF148	zinc finger protein 148	1.E-04	0.050	9.3	2.E-04							✓	✓				✓		✓	✓	
210519_s_at	NQO1	NAD(P)H dehydrogenase, quinone 1	1.E-04	0.050	14.4	0.007			S↓		✓		✓	✓	✓			✓	✓			✓
203549_s_at	LPL	lipoprotein lipase	2.E-04	0.051	16.9	8.E-04				✓	✓	5	✓	✓	✓			✓				
212185_x_at	MT2A	metallothionein 2A	2.E-04	0.051	18.4	6.E-04					✓	2			✓		✓	✓		✓		
221590_s_at	ALDH6A1	aldehyde dehydrogenase 6 family, member A1	2.E-04	0.055	15.8	7.E-04		✓					✓	✓								
202122_s_at	PLIN3	perilipin 3	2.E-04	0.055	10.3	0.001			S↑							✓						
215338_s_at	NKTR	natural killer cell triggering receptor	2.E-04	0.055	10.3	2.E-04						5	✓									
219550_at	ROBO3	roundabout, axon guidance receptor, homolog 3 (Drosophila)	2.E-04	0.057	7.3	0.001			S↓			5		✓							✓	
207543_s_at	P4HA1	prolyl 4-hydroxylase, alpha polypeptide I	2.E-04	0.057	10.7	2.E-04	✓			✓	✓	3,5	✓									
210057_at	SMG1	SMG1 phosphatidylinositol 3-kinase-related kinase	2.E-04	0.057	6.5	4.E-04				✓		3	✓			✓		✓				
214212_x_at	FERMT2	fermitin family member 2	2.E-04	0.057	11.0	5.E-04						5		✓	✓							
203535_at	S100A9	S100 calcium binding protein A9	2.E-04	0.057	18.2	0.003	✓	✓				3	✓		✓	✓	✓	✓	✓	✓		
221527_s_at	PARD3	par-3 family cell polarity regulator	3.E-04	0.062	10.7	0.001						5	✓	✓	✓	✓		✓			✓	
212732_at	MEG3	maternally expressed 3 (non-protein coding)	3.E-04	0.063	21.6	7.E-04																
204846_at	CP	ceruloplasmin (ferroxidase)	3.E-04	0.065	15.4	0.003							✓			✓						
207547_s_at	FAM107A	family with sequence similarity 107, member A	3.E-04	0.067	16.2	0.004			S↓		✓	2					✓					
220856_x_at	NA	NA	3.E-04	0.068	6.8	0.006																
206870_at	PPARA	peroxisome proliferator-activated receptor alpha	4.E-04	0.068	6.4	0.005		✓					✓	✓	✓	✓		✓	✓			
203504_s_at	ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1	4.E-04	0.077	10.7	4.E-04		✓		✓	✓	2,3,5	✓	✓	✓	✓		✓				
209030_s_at	CADM1	cell adhesion molecule 1	5.E-04	0.077	10.0	0.005		✓						✓		✓	✓	✓	✓	✓	✓	
213571_s_at	EIF4E2	eukaryotic translation initiation factor 4E family member 2	5.E-04	0.077	6.6	5.E-04							✓	✓	✓							
212209_at	MED13L	mediator complex subunit 13-like	5.E-04	0.077	7.3	4.E-04						2,3	✓									
203548_s_at	LPL	lipoprotein lipase	5.E-04	0.077	16.4	0.003				✓	✓	5	✓	✓	✓			✓				
204347_at	AK4	adenylate kinase 4	5.E-04	0.077	7.0	7.E-04							✓	✓							✓	
213900_at	FAM189A2	family with sequence similarity 189, member A2	5.E-04	0.077	11.8	0.001	✓	✓				2										
212109_at	HN1L	hematological and neurological expressed 1-like	5.E-04	0.077	5.3	0.005																
218618_s_at	FNDC3B	fibronectin type III domain containing 3B	5.E-04	0.077	8.8	0.001																
204560_at	FKBP5	FK506 binding protein 5	5.E-04	0.077	12.1	0.029						4	✓									
201856_s_at	ZFR	zinc finger RNA binding protein	5.E-04	0.078	8.8	0.002			A↑					✓								
215628_x_at	NA	NA	5.E-04	0.079	9.0	0.008			D↑													
206788_s_at	CBFB	core-binding factor, beta subunit	6.E-04	0.079	7.4	0.050							✓	✓								
212469_at	NIPBL	Nipped-B homolog (Drosophila)	6.E-04	0.079	5.9	0.002							✓	✓			✓	✓			✓	
210400_at	GRIN2C	glutamate receptor, ionotropic, N-methyl D-aspartate 2C	6.E-04	0.079	11.3	0.002			S↓				✓		✓	✓		✓				✓
210068_s_at	AQP4	aquaporin 4	6.E-04	0.080	25.9	0.003		✓				2				✓		✓		✓		
206825_at	OXTR	oxytocin receptor	6.E-04	0.080	10.8	0.001							✓	✓	✓	✓		✓			✓	✓
201999_s_at	DYNLT1	dynein, light chain, Tctex-type 1	6.E-04	0.080	10.2	5.E-04					✓	5	✓	✓	✓	✓					✓	
201468_s_at	NQO1	NAD(P)H dehydrogenase, quinone 1	7.E-04	0.086	10.1	0.014		✓	S↓				✓	✓	✓			✓	✓			✓
206395_at	DGKG	diacylglycerol kinase, gamma 90kDa	7.E-04	0.089	6.9	0.003							✓	✓	✓			✓			✓	
213946_s_at	OBSL1	obscurin-like 1	7.E-04	0.090	7.4	0.004						3		✓							✓	
217730_at	TMBIM1	transmembrane BAX inhibitor motif containing 1	7.E-04	0.091	11.2	0.004					✓		✓	✓	✓				✓			
201666_at	TIMP1	TIMP metallopeptidase inhibitor 1	8.E-04	0.092	17.8	0.002	✓	✓	S↓				✓	✓	✓	✓	✓	✓	✓			
200971_s_at	SERP1	stress-associated endoplasmic reticulum protein 1	8.E-04	0.092	7.8	0.009							✓	✓	✓	✓	✓	✓				

205498_at	GHR	growth hormone receptor	8.E-04	0.092	6.7	0.001			D↑				✓	✓	✓	✓	✓		✓		
211464_x_at	CASP6	caspase 6, apoptosis-related cysteine peptidase	8.E-04	0.093	6.9	5.E-04			A↓ D↑				✓	✓					✓		
206204_at	GRB14	growth factor receptor-bound protein 14	8.E-04	0.093	8.9	0.009						3			✓			✓			
215296_at	CDC42BPA	CDC42 binding protein kinase alpha (DMPK-like)	8.E-04	0.093	7.1	8.E-04															
216814_at	NA	NA	8.E-04	0.093	2.9	0.004															
212208_at	MED13L	mediator complex subunit 13-like	8.E-04	0.094	6.5	0.004							✓								
208990_s_at	HNRNPH3	heterogeneous nuclear ribonucleoprotein H3 (2H9)	9.E-04	0.094	7.1	5.E-04			D↓ AP↑			5	✓	✓							
210588_x_at	HNRNPH3	heterogeneous nuclear ribonucleoprotein H3 (2H9)	9.E-04	0.094	6.3	0.006						5	✓	✓							
201201_at	CSTB	cystatin B (stefin B)	9.E-04	0.096	6.0	0.008		✓	AP↑			5	✓						✓		
211456_x_at	MT1HL1	metallothionein 1H-like 1	9.E-04	0.096	13.4	0.004						2					✓				
202581_at	HSPA1A	heat shock 70kDa protein 1A	9.E-04	0.096	27.2	0.008						3	✓	✓	✓		✓	✓	✓		
209355_s_at	PPAP2B	phosphatidic acid phosphatase type 2B	9.E-04	0.096	24.1	0.007		✓				2	✓	✓	✓						✓
218345_at	TMEM176A	transmembrane protein 176A	9.E-04	0.096	9.4	0.005				✓	✓										
217166_at	PTMAP1	prothymosin, alpha pseudogene 1 (gene sequence 26)	9.E-04	0.096	4.9	0.002															
215592_at	NA	NA	9.E-04	0.096	7.9	0.006			D↑												
220532_s_at	TMEM176B	transmembrane protein 176B	9.E-04	0.098	13.7	0.008				✓	✓	3		✓							
221496_s_at	TOB2	transducer of ERBB2, 2	0.001	0.098	11.5	0.003							✓	✓			✓				
222094_at	LOC101060714	sulfotransferase 1A3/1A4-like	0.001	0.098	5.4	0.008															
201996_s_at	SPEN	spen family transcriptional repressor	0.001	0.098	11.9	5.E-04							✓	✓	✓						✓
209047_at	AQP1	aquaporin 1 (Colton blood group)	0.001	0.099	23.9	0.006						2	✓	✓	✓	✓	✓	✓	✓		✓

(B) Affymetrix ID	Gene symbol	Gene name	P	Q	%	P (no array weights)	Robust regression	Heterogeneity	Potential confounders	Mistry, Gillis, and Pavlidis (2012)	Pérez-Santiago et al (2012)	Agreement with individual cohorts	Biological processes: MP	DP	CC & ST	Transport	Growth / CP	Response to stress	Cell death	IR	NSD	Synaptic transmission
203180_at	ALDH1A3	aldehyde dehydrogenase 1 family, member A3	6.E-08	5.E-04	-12.6	4.E-06						3	✓	✓					✓		✓	
202688_at	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	7.E-08	5.E-04	-23.0	3.E-05				✓	✓	2	✓	✓	✓				✓	✓		
209735_at	ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2 (Junior blood group)	7.E-08	5.E-04	-26.3	1.E-05					✓	2	✓			✓						
217191_x_at	COX6CP1	cytochrome c oxidase subunit VIc pseudogene 1	5.E-07	0.002	-5.6	2.E-06																
201865_x_at	NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)	2.E-06	0.007	-11.5	1.E-05						5	✓	✓	✓				✓			
212987_at	FBXO9	F-box protein 9	4.E-06	0.007	-12.5	8.E-05				✓		1,3	✓	✓	✓			✓		✓		
205336_at	PVALB	parvalbumin	7.E-06	0.011	-21.3	1.E-05	✓				✓											
206414_s_at	ASAP2	ArfGAP with SH3 domain, ankyrin repeat and PH domain 2	8.E-06	0.012	-10.4	0.002							✓									
201310_s_at	NREP	neuronal regeneration related protein	9.E-06	0.013	-10.1	5.E-05				✓		3		✓	✓			✓			✓	
203686_at	MPG	N-methylpurine-DNA glycosylase	1.E-05	0.015	-5.8	9.E-06							✓					✓				
203162_s_at	KATNB1	katanin p80 (WD repeat containing) subunit B 1	1.E-05	0.015	-7.2	7.E-05							✓	✓		✓			✓		✓	
220045_at	NEUROD6	neuronal differentiation 6	1.E-05	0.016	-12.6	5.E-05					✓		✓	✓							✓	
202135_s_at	ACTR1B	ARP1 actin-related protein 1 homolog B, centractin beta (yeast)	2.E-05	0.023	-7.9	2.E-04						3										
203586_s_at	ARL4D	ADP-ribosylation factor-like 4D	3.E-05	0.027	-6.8	9.E-05						3,5	✓		✓	✓						
220741_s_at	PPA2	pyrophosphatase (inorganic) 2	3.E-05	0.029	-7.0	8.E-04				✓		1,3	✓									
217772_s_at	MTCH2	mitochondrial carrier 2	3.E-05	0.029	-8.3	3.E-04										✓			✓			
204218_at	ANAPC15	anaphase promoting complex subunit 15	4.E-05	0.031	-6.0	7.E-04						1										
37005_at	MINOS1-NBL1	MINOS1-NBL1 readthrough	4.E-05	0.031	-7.2	1.E-05		✓						✓	✓						✓	
209019_s_at	PINK1	PTEN induced putative kinase 1	4.E-05	0.034	-7.4	2.E-04		✓				5	✓		✓	✓		✓	✓			✓
208836_at	ATP1B3	ATPase, Na+/K+ transporting, beta 3 polypeptide	5.E-05	0.035	-7.5	3.E-04							✓			✓		✓				
212611_at	DTX4	deltex 4, E3 ubiquitin ligase	5.E-05	0.035	-8.6	4.E-04						3	✓		✓			✓		✓		
205019_s_at	VIPR1	vasoactive intestinal peptide receptor 1	5.E-05	0.035	-7.7	2.E-04						3,5			✓		✓			✓		✓
205508_at	SCN1B	sodium channel, voltage-gated, type I, beta subunit	6.E-05	0.036	-10.8	5.E-05					✓	5	✓	✓	✓	✓	✓		✓		✓	✓
207191_s_at	ISLR	immunoglobulin superfamily containing leucine-rich repeat	6.E-05	0.037	-9.7	0.004			A↓ D↑ S↑													
221494_x_at	EIF3K	eukaryotic translation initiation factor 3, subunit K	6.E-05	0.037	-6.2	0.002							✓									
210638_s_at	FBXO9	F-box protein 9	6.E-05	0.037	-10.8	6.E-04				✓	✓	1,3,5	✓	✓	✓			✓		✓		
217032_at	FOXD4	forkhead box D4	6.E-05	0.037	-6.9	8.E-05							✓									
217934_x_at	STUB1	STIP1 homology and U-box containing protein 1, E3	7.E-05	0.038	-6.0	0.002	✓	✓				5	✓		✓			✓				

		ubiquitin protein ligase																				
205223_at	DEPDC5	DEP domain containing 5	7.E-05	0.038	-5.8	6.E-04			S↑				✓		✓							
221011_s_at	LBH	limb bud and heart development	7.E-05	0.038	-12.1	5.E-04					✓		✓	✓				✓				
206382_s_at	BDNF	brain-derived neurotrophic factor	8.E-05	0.040	-12.0	9.E-05						1,3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
219689_at	SEMA3G	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3G	8.E-05	0.040	-11.3	2.E-04					✓	3		✓								
212716_s_at	EIF3K	eukaryotic translation initiation factor 3, subunit K	8.E-05	0.040	-7.9	0.006						1	✓									
200021_at	CFL1	cofilin 1 (non-muscle)	1.E-04	0.041	-3.0	7.E-04	✓					5	✓	✓	✓	✓		✓	✓	✓	✓	
203273_s_at	TUSC2	tumor suppressor candidate 2	1.E-04	0.041	-5.3	1.E-04			D↓				✓	✓	✓	✓	✓	✓		✓		
203698_s_at	FRZB	frizzled-related protein	1.E-04	0.041	-11.2	0.002		✓	A↑		✓	3		✓	✓		✓		✓		✓	
208478_s_at	BAX	BCL2-associated X protein	1.E-04	0.041	-6.8	8.E-04						5	✓	✓	✓	✓	✓	✓	✓		✓	
219869_s_at	SLC39A8	solute carrier family 39 (zinc transporter), member 8	1.E-04	0.041	-7.7	4.E-04										✓						
213897_s_at	MRPL23	mitochondrial ribosomal protein L23	1.E-04	0.041	-7.7	7.E-04						1	✓									
218105_s_at	MRPL4	mitochondrial ribosomal protein L4	1.E-04	0.044	-5.9	5.E-05						3	✓									
204621_s_at	NR4A2	nuclear receptor subfamily 4, group A, member 2	1.E-04	0.050	-12.3	7.E-04			AP↓		✓	2	✓	✓	✓			✓	✓		✓	
201900_s_at	AKR1A1	aldo-keto reductase family 1, member A1 (aldehyde reductase)	2.E-04	0.051	-6.3	0.001	✓		AP↓				✓									
201694_s_at	EGR1	early growth response 1	2.E-04	0.051	-15.7	0.002				✓	✓	3	✓	✓	✓		✓	✓	✓	✓	✓	✓
206209_s_at	CA4	carbonic anhydrase IV	2.E-04	0.051	-10.4	5.E-04			A↓ AP↓	✓	✓	3	✓	✓		✓						
203286_at	RNF44	ring finger protein 44	2.E-04	0.051	-6.0	3.E-04																
205061_s_at	EXOSC9	exosome component 9	2.E-04	0.051	-5.0	3.E-04				✓		3	✓				✓			✓		
205139_s_at	UST	uronyl-2-sulfotransferase	2.E-04	0.052	-10.0	0.009						4	✓									
201740_at	NDUFS3	NADH dehydrogenase (ubiquinone) Fe-S protein 3, 30kDa (NADH-coenzyme Q reductase)	2.E-04	0.055	-5.9	0.001			D↓			3	✓	✓	✓		✓		✓		✓	
218760_at	COQ6	coenzyme Q6 monooxygenase	2.E-04	0.055	-4.8	0.002			D↑				✓									
208827_at	PSMB6	proteasome (prosome, macropain) subunit, beta type, 6	2.E-04	0.055	-7.2	6.E-04			S↑			1	✓		✓			✓	✓			
202596_at	ENSA	endosulfine alpha	2.E-04	0.055	-9.4	0.001				✓			✓		✓	✓						
219310_at	SYNDIG1	synapse differentiation inducing 1	2.E-04	0.055	-8.5	0.003						1		✓	✓	✓					✓	✓
207767_s_at	EGR4	early growth response 4	2.E-04	0.055	-12.1	0.001	✓				✓		✓				✓		✓			
32502_at	GDPD5	glycerophosphodiester phosphodiesterase domain containing 5	2.E-04	0.057	-5.2	9.E-04							✓	✓	✓						✓	
212646_at	RFTN1	raftlin, lipid raft linker 1	2.E-04	0.057	-8.5	2.E-04			S↓	✓					✓	✓	✓	✓		✓		
200657_at	SLC25A5	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 5	2.E-04	0.057	-7.2	0.002						1,3	✓		✓	✓	✓		✓			
209803_s_at	PHLDA2	pleckstrin homology-like domain, family A, member 2	2.E-04	0.057	-11.3	8.E-04							✓	✓					✓			
212471_at	AVL9	AVL9 homolog (S. cerevisiase)	2.E-04	0.057	-5.4	2.E-04																
215518_at	STXBP5L	syntaxin binding protein 5-like	2.E-04	0.058	-9.3	0.003									✓	✓						
213125_at	OLFML2B	olfactomedin-like 2B	2.E-04	0.058	-7.8	0.012			AP↓													
214674_at	USP19	ubiquitin specific peptidase 19	3.E-04	0.059	-5.6	0.002			A↓ D↑	✓		3	✓	✓				✓				
204789_at	FMNL1	formin-like 1	3.E-04	0.062	-6.8	0.004								✓								
202687_s_at	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	3.E-04	0.062	-9.8	0.005							✓	✓	✓				✓	✓		
215639_at	SH2D3C	SH2 domain containing 3C	3.E-04	0.062	-5.5	0.002																
205249_at	EGR2	early growth response 2	3.E-04	0.062	-11.6	3.E-04	✓	✓	S↑ A↓ D↑ AP↓			2,3										
203720_s_at	ERCC1	excision repair cross-complementation group 1	3.E-04	0.062	-5.2	0.001		✓					✓	✓			✓	✓		✓		

203769_s_at	STS	steroid sulfatase (microsomal), isozyme S	3.E-04	0.065	-8.3	6.E-04				✓			✓	✓							
219032_x_at	OPN3	opsin 3	3.E-04	0.065	-10.8	0.007			S↓	✓		5	✓		✓						
209192_x_at	KAT5	K(lysine) acetyltransferase 5	3.E-04	0.065	-4.9	6.E-04							✓		✓		✓	✓			
205483_s_at	ISG15	ISG15 ubiquitin-like modifier	3.E-04	0.067	-10.3	3.E-04							✓	✓	✓			✓		✓	
210910_s_at	POMZP3	POM121 and ZP3 fusion	3.E-04	0.068	-15.0	0.007	✓						✓	✓	✓	✓	✓	✓		✓	
202698_x_at	COX4I1	cytochrome c oxidase subunit IV isoform 1	3.E-04	0.068	-4.2	9.E-04							✓			✓					
210501_x_at	EIF3K	eukaryotic translation initiation factor 3, subunit K	4.E-04	0.068	-5.3	0.006		✓				1	✓								
209993_at	ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	4.E-04	0.068	-11.1	0.006							✓			✓	✓				
203719_at	ERCC1	excision repair cross-complementation group 1	4.E-04	0.068	-4.8	0.003				✓			✓					✓			
203697_at	FRZB	frizzled-related protein	4.E-04	0.069	-13.8	0.014					✓	3		✓	✓		✓		✓		✓
216248_s_at	NR4A2	nuclear receptor subfamily 4, group A, member 2	4.E-04	0.069	-14.3	0.011					✓		✓	✓	✓			✓	✓		✓
208855_s_at	STK24	serine/threonine kinase 24	4.E-04	0.069	-5.4	0.006	✓						✓	✓	✓			✓	✓		✓
202746_at	ITM2A	integral membrane protein 2A	4.E-04	0.070	-13.0	0.011															
213316_at	KIAA1462	KIAA1462	4.E-04	0.071	-9.1	0.010						1									
201002_s_at	TMEM189-UBE2V1	TMEM189-UBE2V1 readthrough	4.E-04	0.072	-6.2	0.008						3	✓	✓	✓			✓		✓	
217757_at	A2M	alpha-2-macroglobulin	4.E-04	0.074	-13.5	0.007							✓	✓	✓	✓		✓		✓	
204304_s_at	PROM1	prominin 1	4.E-04	0.077	-8.8	0.038					✓			✓							✓
213791_at	PENK	proenkephalin	5.E-04	0.077	-10.7	0.002		✓			✓	2			✓			✓		✓	
207988_s_at	ARPC2	actin related protein 2/3 complex, subunit 2, 34kDa	5.E-04	0.077	-6.6	0.002								✓	✓	✓		✓		✓	
204622_x_at	NR4A2	nuclear receptor subfamily 4, group A, member 2	5.E-04	0.077	-12.5	1.E-03	✓					2	✓	✓	✓			✓	✓		✓
221483_s_at	ARPP19	cAMP-regulated phosphoprotein, 19kDa	5.E-04	0.077	-5.3	0.003						3	✓			✓					
207515_s_at	POLR1C	polymerase (RNA) I polypeptide C, 30kDa	5.E-04	0.077	-5.1	0.003							✓					✓		✓	
213385_at	CHN2	chimerin 2	5.E-04	0.077	-10.7	0.001							✓		✓						
205114_s_at	CCL3	chemokine (C-C motif) ligand 3	5.E-04	0.077	-10.9	2.E-04	✓			✓		3	✓	✓	✓	✓	✓	✓	✓	✓	✓
210605_s_at	MFGE8	milk fat globule-EGF factor 8 protein	5.E-04	0.077	-6.2	0.002								✓		✓					
204286_s_at	PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1	5.E-04	0.077	-7.7	0.001							✓		✓	✓	✓	✓	✓	✓	
201044_x_at	DUSP1	dual specificity phosphatase 1	5.E-04	0.078	-8.8	6.E-04							✓	✓	✓			✓	✓		
203272_s_at	TUSC2	tumor suppressor candidate 2	5.E-04	0.078	-7.2	4.E-04						3	✓	✓	✓	✓	✓	✓		✓	
219123_at	ZNF232	zinc finger protein 232	5.E-04	0.078	-5.8	0.015							✓								
211671_s_at	NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)	5.E-04	0.079	-10.1	0.002							✓	✓	✓				✓		
214112_s_at	CXorf40A	chromosome X open reading frame 40A	5.E-04	0.079	-5.9	0.003															
218651_s_at	LARP6	La ribonucleoprotein domain family, member 6	5.E-04	0.079	-7.8	0.002							✓								
203851_at	IGFBP6	insulin-like growth factor binding protein 6	5.E-04	0.079	-7.1	6.E-04				✓			✓		✓		✓				
203367_at	DUSP14	dual specificity phosphatase 14	6.E-04	0.079	-7.3	0.006	✓					5	✓		✓						
201851_at	SH3GL1	SH3-domain GRB2-like 1	6.E-04	0.079	-4.6	0.001								✓	✓	✓					✓
210720_s_at	NECAB3	N-terminal EF-hand calcium binding protein 3	6.E-04	0.079	-6.4	5.E-04				✓			✓			✓					
213366_x_at	ATP5C1	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	6.E-04	0.079	-5.7	0.007			S↑	✓		3	✓			✓					
44065_at	RITA1	RBPJ interacting and tubulin associated 1	6.E-04	0.080	-6.1	0.008						1	✓	✓	✓	✓					✓
204794_at	DUSP2	dual specificity phosphatase 2	6.E-04	0.081	-6.7	0.002	✓						✓	✓	✓						
201199_s_at	PSMD1	proteasome (prosome, macropain) 26S subunit, non-ATPase, 1	6.E-04	0.083	-5.9	0.010							✓		✓			✓	✓		
210033_s_at	SPAG6	sperm associated antigen 6	6.E-04	0.083	-6.8	0.001								✓							
205381_at	LRRC17	leucine rich repeat containing 17	6.E-04	0.084	-7.3	0.010				✓		3		✓							
203258_at	DRAP1	DR1-associated protein 1 (negative cofactor 2 alpha)	6.E-04	0.085	-5.3	0.003							✓								
200925_at	COX6A1	cytochrome c oxidase subunit VIa polypeptide 1	7.E-04	0.085	-4.3	0.010							✓			✓					

209267_s_at	SLC39A8	solute carrier family 39 (zinc transporter), member 8	7.E-04	0.085	-8.4	0.011										✓						
205984_at	CRHBP	corticotropin releasing hormone binding protein	7.E-04	0.086	-10.8	0.002		✓	S↑ AP↓			1,2	✓		✓	✓		✓		✓		✓
200948_at	MLF2	myeloid leukemia factor 2	7.E-04	0.087	-5.9	0.002						5						✓		✓		
49077_at	PPME1	protein phosphatase methylesterase 1	7.E-04	0.087	-5.3	0.004							✓									
201677_at	HMCES	5-hydroxymethylcytosine (hmC) binding, ES cell-specific	7.E-04	0.090	-7.1	0.014							✓									
40472_at	LPCAT4	lysophosphatidylcholine acyltransferase 4	7.E-04	0.090	-7.7	0.004						5	✓									
214365_at	TPM3	tropomyosin 3	7.E-04	0.091	-8.1	0.003																
206850_at	RASL10A	RAS-like, family 10, member A	7.E-04	0.091	-8.5	0.011		✓	A↓				✓		✓	✓						
203872_at	ACTA1	actin, alpha 1, skeletal muscle	8.E-04	0.092	-5.6	0.002			A↑					✓			✓					
209840_s_at	LRRN3	leucine rich repeat neuronal 3	8.E-04	0.092	-9.4	0.004																
217405_x_at	GPLD1	glycosylphosphatidylinositol specific phospholipase D1	8.E-04	0.093	-5.0	0.001		✓					✓	✓	✓	✓	✓		✓	✓		
208165_s_at	PRSS16	protease, serine, 16 (thymus)	8.E-04	0.093	-5.4	5.E-04							✓									
219825_at	CYP26B1	cytochrome P450, family 26, subfamily B, polypeptide 1	8.E-04	0.093	-13.9	0.001			D↓	✓	✓	3	✓	✓	✓				✓			
220482_s_at	SERGEF	secretion regulating guanine nucleotide exchange factor	8.E-04	0.093	-6.2	0.005	✓						✓		✓	✓						
217992_s_at	EFHD2	EF-hand domain family, member D2	8.E-04	0.093	-8.4	0.002		✓				5										
213851_at	TMEM110	transmembrane protein 110	8.E-04	0.093	-5.4	0.001																
213041_s_at	ATP5D	ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit	8.E-04	0.093	-6.7	0.004						3	✓			✓						
215716_s_at	ATP2B1	ATPase, Ca++ transporting, plasma membrane 1	8.E-04	0.094	-11.5	0.009						1,5	✓			✓		✓				
203694_s_at	DHX16	DEAH (Asp-Glu-Ala-His) box polypeptide 16	9.E-04	0.096	-5.2	0.001							✓									
209270_at	LAMB3	laminin, beta 3	9.E-04	0.096	-7.8	6.E-04	✓	✓	S↑ AP↓					✓								
213750_at	RSL1D1	ribosomal L1 domain containing 1	9.E-04	0.096	-7.4	0.025		✓						✓				✓	✓			
211725_s_at	BID	BH3 interacting domain death agonist	9.E-04	0.096	-7.4	0.001		✓					✓	✓	✓	✓	✓	✓	✓		✓	
203831_at	R3HDM2	R3H domain containing 2	9.E-04	0.098	-6.2	0.014			S↓			3,5										
218262_at	RMND5B	required for meiotic nuclear division 5 homolog B (S. cerevisiae)	9.E-04	0.098	-4.6	0.014				✓		3										
206258_at	ST8SIA5	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 5	0.001	0.098	-6.7	0.002							✓									
212252_at	CAMKK2	calcium/calmodulin-dependent protein kinase kinase 2, beta	0.001	0.098	-8.2	7.E-05				✓		5	✓		✓							
203541_s_at	KLF9	Kruppel-like factor 9	0.001	0.098	-5.5	0.003							✓	✓	✓							
209699_x_at	AKR1C2	aldo-keto reductase family 1, member C2	0.001	0.098	-7.2	8.E-04				✓			✓	✓	✓	✓	✓					
218097_s_at	CUEDC2	CUE domain containing 2	0.001	0.098	-5.0	0.003																
206375_s_at	HSPB3	heat shock 27kDa protein 3	0.001	0.098	-8.8	0.001												✓	✓			
201328_at	ETS2	v-ets avian erythroblastosis virus E26 oncogene homolog 2	0.001	0.099	-8.5	0.005				✓			✓	✓								
203349_s_at	ETV5	ets variant 5	0.001	0.099	-10.2	0.013			A↓	✓	✓	1,2,3	✓	✓	✓		✓	✓			✓	✓
205413_at	MPPED2	metallophosphoesterase domain containing 2	0.001	0.100	-10.0	0.010				✓	✓	3	✓	✓							✓	

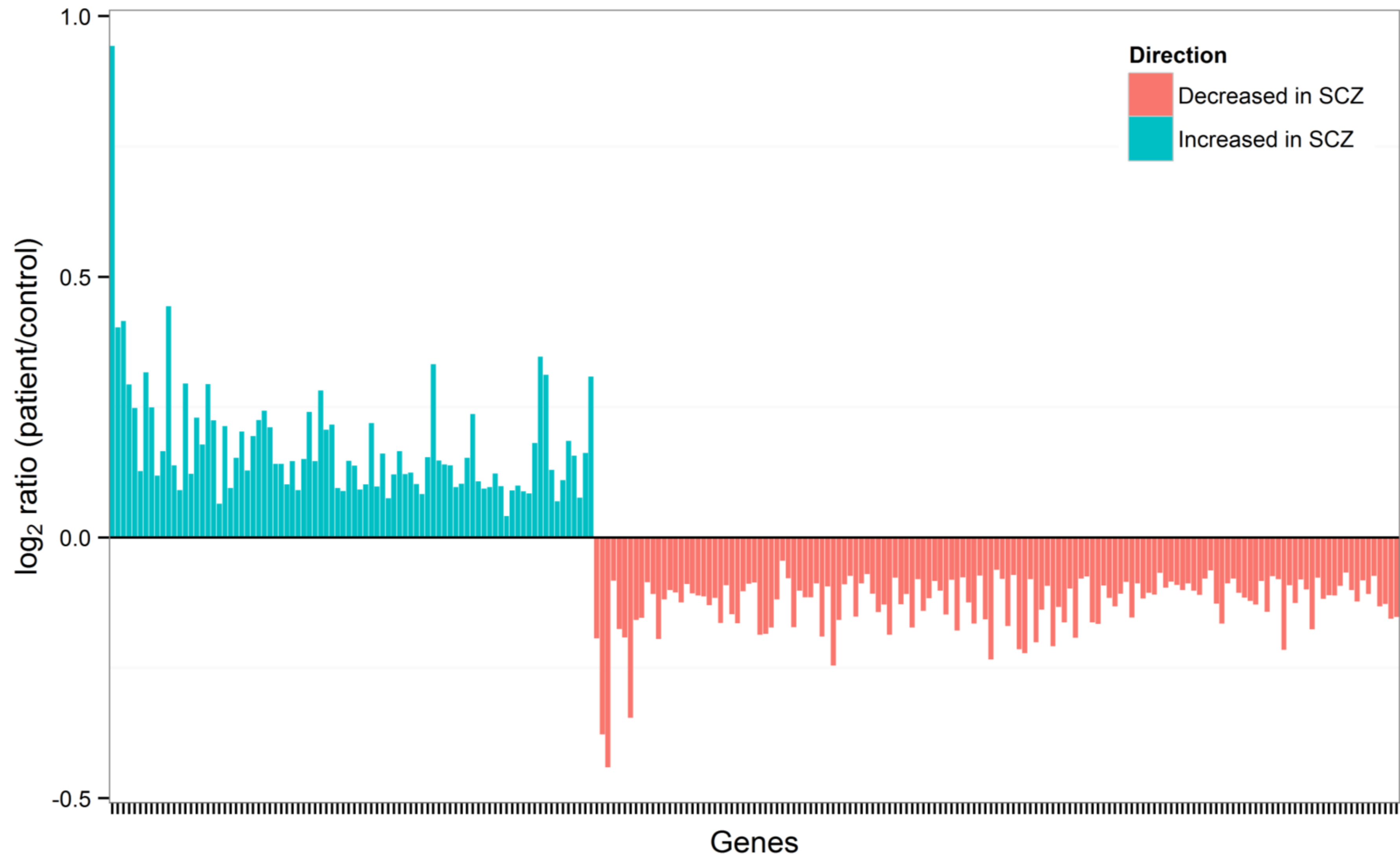


Figure 6.3. Probe sets with significantly different intensity between schizophrenia patients and controls in the PFC. Probes are listed in the same order as in **Table 6.2**. Probe set intensities in blue were higher in schizophrenia and probe set intensities in pink were lower. The $\log_2$ ratios of PFC gene expression in schizophrenia patients compared to controls were coefficients from the linear regression. **Abbreviation:** SCZ (schizophrenia).

Table 6.3. **Results of analysis with and without the use of array weights and robust regression in the presence of regression outlier(s).**

Case	Analysis using *limma*		Number of significant probe sets (q<0.1)	Overlap with Case 1
	Array weights	Robust regression		
1	✓	✓	229	
2	✓		207	205
3		✓	27	26
4			17	17

Table 6.4. Enriched GO BP terms for differentially expressed genes in schizophrenia. Criteria for enrichment are listed in **Section 6.2.6.** Genes in **red** have increased expression in schizophrenia and genes in **green** have reduced expression in schizophrenia.

GO BP ID	Term	Expected significant genes	Number of significant genes	Annotated genes	P-value	Significant genes
GO:1990267	response to transition metal nanoparticle	1.2	9	74	3.E-06	ANK2, AQP1, MT1E, MT1M, MT1X, MT2A, ATP5D, S100A8, S100A9
GO:0010038	response to metal ion	3.8	13	232	1.E-04	CRHBP, DUSP1, GPLD1, AQP1, MT1E, MT1M, MT1X, MT2A, ATP5D, ACTA1, S100A8, S100A9, ABCG2
GO:0071383	cellular response to steroid hormone stimulus	1.7	8	107	3.E-04	KAT5, CRHBP, EGR1, NR3C1, AQP1, AQP4, NR4A2, PPARA
GO:0033993	response to lipid	10.0	22	613	4.E-04	KAT5, CRHBP, AKR1C2, NQO1, DUSP1, ABCA1, EGR1, GHR, GPLD1, NR3C1, AQP1, AQP4, NR4A2, OXTR, PPARA, CYP26B1, ACTA1, BDNF, S100A8, S100A9, CA4, ABCG2
GO:0090200	positive regulation of release of cytochrome c from mitochondria	0.4	4	22	4.E-04	PMAIP1, BAX, PINK1, TNFSF10
GO:0051239	regulation of multicellular organismal process	29.1	47	1790	4.E-04	LRRC17, KATNB1, KAT5, TOB2, USP19, TUSC2, CRHBP, EGR1, EGR2, ETV5, SPEN, DTX4, OBSL1, CADM1, FRZB, NIPBL, GHR, SERP1, GPLD1, ANK2, TMEM176B, HSPB1, AQP1, NR4A2, OXTR, TMEM176A, PARD3, CYP26B1, BAX, BDNF, SCN1B, TMBIM1, FNDC3B, DYNLT1, TIMP1, PHLDA2, SYNDIG1, GDPD5, LBH, STK24, PPAP2B, PROM1, NOL3, DHRS3, NREP, POLR1C, ISG15
GO:0043525	positive regulation of neuron apoptotic process	0.7	5	40	4.E-04	NQO1, EGR1, NR3C1, PMAIP1, BAX
GO:0071320	cellular response to cAMP	0.7	5	40	4.E-04	CRHBP, EGR1, EGR2, EGR4, AQP1
GO:0032355	response to estradiol	1.4	7	89	6.E-04	KAT5, CRHBP, NQO1, DUSP1, GHR, AQP4, OXTR
GO:0014070	response to organic cyclic compound	10.4	22	640	7.E-04	KAT5, CRHBP, NQO1, DUSP1, ABCA1, EGR1, EGR2, EGR4, RFTN1, GHR, GPLD1, NR3C1, AQP1, AQP4, NR4A2, OXTR, PMAIP1, PPARA, ACTA1, BDNF, CA4, ABCG2
GO:0022904	respiratory electron transport chain	1.5	7	92	7.E-04	COX4I1, COX6A1, NDUFS3, ATP5C1, ATP5D, BDNF, PINK1
GO:0097306	cellular response to alcohol	1.1	6	67	7.E-04	KAT5, CRHBP, ABCA1, GPLD1, AQP1, AQP4

Fifty of the differentially expressed genes in schizophrenia were associated with suicide or lifetime drug, alcohol, or antipsychotic use in the SMRI cohorts ($p<0.05$). Expression for 29 of these genes was associated with potential confounders in the same direction as schizophrenia diagnosis. Increased expression of C4B, NOL3, HNRNPH3, S100A8, and CSTB and decreased expression of CRHBP, LAMB3, EGR2, OLFML2B, CA4, AKR1A1, and NR4A2 occurred with increased lifetime antipsychotic use and was consistent with expression in schizophrenia patients. Both schizophrenia diagnosis and moderate/severe lifetime alcohol or drug use were also related to decreased expression of ETV5, RASL10A, EGR2, USP19, CA4, ISLR, TUSC2, CYP26B1, and NDUFS3 and increased expression of ZFR, GHR, CASP6, and two other probe sets (Affymetrix IDs 215628_x_at and 215592_at). Suicide as cause of death was related to increased expression of PLIN3 and decreased expression of RFTN1 and OPN3, also consistent with expression in schizophrenia.

A comparison of microarray and serum findings (**Chapter 5**) revealed that both PFC gene expression and serum protein concentration of tissue inhibitor of metalloproteinases 1 (TIMP-1) was increased in schizophrenia. Conversely, α2-macroglobulin protein concentration was increased in serum, but gene expression was decreased in the PFC. No other overlap was found with differentially expressed PFC genes using $q<0.1$. However, relaxing the *q*-value cut-off criteria, an additional six genes overlapped with serum results ($p<0.05$) and were in the same direction. Both PFC gene expression and serum protein concentration of ferritin, carcinoembryonic antigen (CEA), cancer antigen 19-9 (CA 19-9), EN-RAGE, hepatocyte growth factor (HGF), and α1-antitrypsin were increased in schizophrenia (**Table 6.5**). On the other hand, AXL receptor tyrosine kinase was decreased in schizophrenia patient serum, but gene expression was increased in the PFC.

Table 6.5. Comparison of PFC gene expression findings with serum findings for first onset, antipsychotic naive schizophrenia patients compared to controls in Chapter 5.

Serum (Chapter 5)				PFC			
Analytes	**UniProt/ PubChem identifier**	**Q-value**	**%**	**Microarray probe set**	**P-value**	**Q-value**	**%**
Ferritin	P02792, P02794	1.E-10	125.9	213187_x_at / 214211_at	0.002 / 0.020	0.14 / 0.30	8.1 / 6.0
MIF	P14174	8.E-10	99.6	217871_s_at	0.52	0.86	-1.5
Haptoglobin	P00738	2.E-06	59.6	206697_s_at	0.90	0.98	0.2
CEA	P06731	2.E-06	57.7	217291_at	0.042	0.39	2.8
PP	P01298	0.026	42.0	210670_at	0.64	0.90	-0.61
Cortisol	5754	5.E-08	37.7	Not measured			
FSH	P01225, P01215	0.004	37.4	214489_at / 204637_at	0.43 / 0.036	0.83 / 0.38	-1.4 / 4.1
LH	P01229, P01215	0.006	31.8	214471_x_at / 204637_at	0.57 / 0.036	0.88 / 0.38	-1.2 / 4.1
CA 19-9	Q9BXJ9	0.043	30.4	219158_s_at	0.032	0.36	5.3
EN-RAGE	P80511	0.043	26.5	205863_at	0.009	0.22	9.0
Testosterone	6013	4.E-05	24.4	Not measured			
CgA	P10645	0.027	22.7	204697_s_at	0.086	0.51	-4.1
IL-8	P10145	0.011	20.6	202859_x_at	0.21	0.68	-4.3
G-CSF	P09919	0.027	20.3	210228_at	0.81	0.96	-0.3
Progesterone	5994	0.010	19.5	Not measured			
HGF	P14210	0.037	17.3	209960_at	0.002	0.13	7.8
RANTES	P13501	0.030	16.8	204655_at	0.36	0.79	-2.0
Eotaxin-1	P51671	0.038	16.7	210133_at	0.83	0.96	-0.4
A2Macro	P01023	8.E-07	16.2	217757_at	4E-04	0.074	-13.5
IGFBP-2	P18065	0.038	14.7	202718_at	0.23	0.70	-3.4
TN-C	P24821	0.037	14.1	216005_at	0.66	0.91	-0.75
PDGF	P01127	0.030	13.5	216055_at	0.20	0.67	1.9
IL-10	P22301	4.E-05	12.6	207433_at	0.61	0.89	-0.6
IL-18	Q14116	0.038	11.0	206295_at	0.44	0.83	-1.1
HCC-4	O15467	0.060	10.8	207354_at	0.65	0.91	0.8
MDC	O00626	0.027	9.9	207861_at	0.88	0.97	-0.2
SAP	P02743	0.038	9.4	206350_at	0.35	0.78	-1.2
IL-15	P40933	0.030	9.2	205992_s_at	0.34	0.78	1.4
AAT	P01009	0.012	8.2	202833_s_at	0.005	0.18	12.0
TNFR2	P20333	0.038	7.6	203508_at	0.76	0.94	0.6
TBG	P05543	0.032	7.5	206386_at	0.85	0.97	0.3
TIMP-1	P01033	0.013	7.0	201666_at	8E-04	0.092	17.8
C3	P01024	0.010	7.0	217767_at	0.76	0.94	2.0
Apo H	P02749	0.018	6.8	205216_s_at	0.75	0.94	0.5
AXL	P30530	0.098	-6.9	202685_s_at	0.027	0.34	3.5
PAP	P15309	0.060	-8.8	204393_s_at	0.62	0.90	0.6
Resistin	Q9HD89	0.074	-9.9	220570_at	0.68	0.92	0.6
Apo A-I	P02647	0.026	-10.6	217073_x_at	0.42	0.82	-1.5
Apo-CIII	P02656	0.013	-11.6	205820_s_at	0.65	0.91	-0.6
TSH	P01222, P01215	0.071	-12.6	214529_at / 204637_at	0.56 / 0.036	0.88 / 0.38	-0.8 / 4.1
Sortilin	Q99523	3.E-04	-15.1	212797_at	0.75	0.94	-1.1
RAGE	Q15109	0.027	-15.2	217046_s_at	0.51	0.86	-1.2
CK-MB	P12277, P06732	0.007	-24.3	200884_at / 204810_s_at	0.66 / 0.29	0.91 / 0.75	-0.7 / -1.8
CD40L	P29965	0.030	-24.6	207892_at	0.30	0.75	-1.6
Leptin	P41159	0.042	-25.6	207092_at	0.71	0.92	0.6

6.3.2 Sex-biased gene expression

Thirty-three genes [46 probe sets] were differentially expressed in control male compared to control female PFC (**Table 6.6** and **Figure 6.4**) in the linear model. Fourteen of the probe sets had higher intensity in females compared to males, while the remaining 32 were higher in males. Eighteen (39%) of these probe sets showed evidence of heterogeneity between cohorts. The majority of sex-biased genes were located on the X- or Y- chromosomes. Sex biased genes were associated with over 1000 GO BP terms and enriched for four of these, including histone demethylation and translation initiation (**Table 6.7**). Genes involved in these enriched functions comprised 21% of the sex-biased genes. Sex differences in serum protein concentration found in **Chapter 3** were compared to sex differences in PFC gene expression, but this revealed limited overlap. Adiponectin (207175_at) PFC gene expression and serum concentration were both higher in females (p=0.046, q=0.72). Cancer antigen 15-3 (207847_s_at; p=0.027, q=0.68), neuronal cell adhesion molecule (204105_s_at; p=0.037, q=0.70), and monocyte chemotactic protein 2 (214038_at; p=0.003, q=0.45) were also consistent with serum results and were higher in males.

6.3.3 Sex-schizophrenia diagnosis interactions in PFC gene expression

No probe set intensities showed significant sex-diagnosis interactions in the linear model with q<0.1. Probe sets having sex-diagnosis interactions with p<0.01 were examined (**Table 6.8** and **Figure 6.5**), all of which had q=1.0 with the exception of HSPA6 (q=0.43). There were 105 genes [107 probe sets] meeting this criteria, 5 of which showed male-specific differences in probe set intensity in schizophrenia, 41 of which showed female-specific differences, and the remaining 61 of which showed qualitative interactions. Eleven of these (10%) showed evidence of heterogeneity between cohorts. These 105 genes were associated with over 2000 GO BP terms, but were not significantly enriched for any of these. There was no overlap between them and the serum proteins found in **Chapter 5** to have a significant sex-diagnosis interaction in concentration (or with genes with p<0.05). None of the probe sets showing different intensity in schizophrenia patients in **Table 6.2** (q<0.1) had a sex-diagnosis interaction with p<0.01.

Table 6.6. **Probe sets with significantly different intensity between male and female control PFC.** In **(A)** probe set intensity is higher in males, while in **(B)** probe set intensity is higher in females. The percentage impact on serum molecular concentrations of male (A) and female (B) sex (%) from the linear regression is reported and use of robust regression is marked for each probe set. GO terms were used to assess participation of a gene in each biological process. **Abbreviations:** P (*p*-value); Q (*q*-value); A (lifetime alcohol use); AP (lifetime antipsychotic use); D (lifetime drug use); S (suicide); MP (metabolic process); DP (developmental process); CC & ST (cell communication and signal transduction); IR (defense/immune/inflammatory response); CP (cell proliferation); NSD (nervous system development). **GO terms (left to right):** GO:0008152; GO:0032502; GO:0007154 and GO:0007165; GO:0006810; GO:0040007 and GO:0008283; GO:0006950; GO:0008219; GO:0006952 and GO:0006955 and GO:0006954; GO:0007399; GO:0007268.

(A)										Biological functions									
Affymetrix ID	**Gene symbol**	**Gene name**	**%**	**P**	**Q**	**Chromosome**	**Band**	**Robust regression**	**Heterogeneity**	**MP**	**DP**	**CC & ST**	**Transport**	**Growth / CP**	**Response to stress**	**Cell death**	**IR**	**NSD**	**Synaptic transmission**
201909_at	RPS4Y1	ribosomal protein S4, Y-linked 1	924.6	2.E-73	4.E-69	Y	p11.2	✓	✓	✓	✓		✓						
205000_at	DDX3Y	DEAD (Asp-Glu-Ala-Asp) box helicase 3, Y-linked	707.1	1.E-67	1.E-63	Y	q11.221	✓	✓	✓									
206700_s_at	KDM5D	lysine (K)-specific demethylase 5D	444.0	9.E-57	4.E-53	Y	q11.223	✓	✓	✓									
214983_at	TTTY15	testis-specific transcript, Y-linked 15 (non-protein coding)	113.4	7.E-56	2.E-52	Y	q11.221	✓											
206624_at	USP9Y	ubiquitin specific peptidase 9, Y-linked	288.4	8.E-46	2.E-42	Y	q11.221	✓	✓	✓		✓							
207703_at	NLGN4Y	neuroligin 4, Y-linked	142.1	8.E-42	2.E-38	Y	q11.221	✓	✓	✓	✓							✓	
204409_s_at	EIF1AY	eukaryotic translation initiation factor 1A, Y-linked	172.4	3.E-37	8.E-34	Y	q11.223	✓	✓	✓									
205001_s_at	DDX3Y	DEAD (Asp-Glu-Ala-Asp) box helicase 3, Y-linked	103.9	1.E-35	2.E-32	Y	q11.221	✓	✓	✓									
207063_at	TTTY14	testis-specific transcript, Y-linked 14 (non-protein coding)	78.3	2.E-34	5.E-31	Y	q11.222	✓	✓										
211149_at	UTY	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	31.2	5.E-26	1.E-22	Y	q11.221	✓	✓	✓									
204410_at	EIF1AY	eukaryotic translation initiation factor 1A, Y-linked	84.8	6.E-24	1.E-20	Y	q11.223	✓	✓	✓									
36554_at	ASMTL	acetylserotonin O-methyltransferase-like	42.2	5.E-15	8.E-12	X/Y	p22.33/ p11.3		✓	✓									
209394_at	ASMTL	acetylserotonin O-methyltransferase-like	30.0	5.E-13	8.E-10	X/Y	p22.33/ p11.3			✓									

218951_s_at	PLCXD1	phosphatidylinositol-specific phospholipase C, X domain containing 1	40.6	7.E-13	1.E-09	X/Y	p22.33/ p11.32			✓									
36553_at	ASMTL	acetylserotonin O-methyltransferase-like	35.8	7.E-13	1.E-09	X/Y	p22.33/ p11.3			✓									
210292_s_at	PCDH11X PCDH11Y	protocadherin 11 X/Y-linked	59.8	1.E-12	1.E-09	X/Y	q21.31/ p11.2			✓									
214131_at	TXLNGY	taxilin gamma pseudogene, Y-linked	40.9	9.E-12	1.E-08	Y	q11.222		✓										
203043_at	ZBED1	zinc finger, BED-type containing 1	19.0	5.E-10	5.E-07	X/Y	p22.33/ p11		✓	✓									
201029_s_at	CD99	CD99 molecule	45.7	6.E-10	6.E-07	X/Y	p22.33/ p11.3												
203314_at	GTPBP6	GTP binding protein 6 (putative)	18.2	4.E-09	4.E-06	X/Y	p22.33/ p11.32												
201028_s_at	CD99	CD99 molecule	43.7	1.E-08	1.E-05	X/Y	p22.33/ p11.3												
210322_x_at	UTY	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	13.8	6.E-08	5.E-05	Y	q11.221		✓	✓									
266_s_at	CD24	CD24 molecule	26.5	9.E-08	7.E-05	6	q21	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
207246_at	ZFY	zinc finger protein, Y-linked	10.6	1.E-07	9.E-05	Y	p11.2			✓									
216379_x_at	CD24	CD24 molecule	36.5	2.E-07	2.E-04	6	q21			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
209771_x_at	CD24	CD24 molecule	33.5	1.E-06	0.001	6	q21		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
211227_s_at	PCDH11X PCDH11Y	protocadherin 11 X/Y-linked	10.5	2.E-06	0.001	X/Y	q21.31/ p11.2												
208748_s_at	FLOT1	flotillin 1	15.9	4.E-05	0.023	6	p21.33				✓	✓						✓	
208651_x_at	CD24	CD24 molecule	16.3	5.E-05	0.026	6	q21			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
205378_s_at	ACHE	acetylcholinesterase (Yt blood group)	15.9	9.E-05	0.045	7	q22.1			✓	✓	✓	✓	✓	✓			✓	✓
203290_at	HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	10.9	9.E-05	0.046			✓											
218784_s_at	SAYSD1	SAYSVFN motif domain containing 1	9.2	1.E-04	0.067	6	p21.2												

(B)

										Biological functions									
Affymetrix ID	Gene symbol	Gene name	%	P	Q	Chromosome	Band	Robust regression	Heterogeneity	MP	DP	CC & ST	Transport	Growth / CP	Response to stress	Cell death	IR	NSD	Synaptic transmission
221728_x_at	XIST	X inactive specific transcript (non-protein coding)	2278.2	1.E-64	9.E-61	X	q13.2	✓	✓										
214218_s_at	XIST	X inactive specific transcript (non-protein coding)	1480.4	2.E-60	1.E-56	X	q13.2	✓	✓										
203974_at	HDHD1	haloacid dehalogenase-like hydrolase domain containing 1	17.9	1.E-08	1.E-05	X	p22.31			✓									
201210_at	DDX3X	DEAD (Asp-Glu-Ala-Asp) box helicase 3, X-linked	21.9	1.E-08	1.E-05	X	p11.4			✓		✓		✓	✓	✓	✓		
203992_s_at	KDM6A	lysine (K)-specific demethylase 6A	26.0	3.E-08	3.E-05	X	p11.3			✓	✓	✓		✓				✓	
200933_x_at	RPS4X	ribosomal protein S4, X-linked	10.8	7.E-07	5.E-04	X	q13.1			✓	✓		✓	✓					
213347_x_at	RPS4X	ribosomal protein S4, X-linked	11.7	2.E-06	0.001	X	q13.1			✓	✓		✓	✓					
212515_s_at	DDX3X	DEAD (Asp-Glu-Ala-Asp) box helicase 3, X-linked	25.9	3.E-06	0.002	X/8	p11.4/p21.2			✓		✓		✓	✓	✓	✓		
209740_s_at	PNPLA4	patatin-like phospholipase domain containing 4	16.0	3.E-06	0.002	X	p22.31			✓									
201100_s_at	USP9X	ubiquitin specific peptidase 9, X-linked	9.8	4.E-05	0.022	X	p11.4			✓	✓	✓		✓				✓	
204121_at	GADD45G	growth arrest and DNA-damage-inducible, gamma	14.7	4.E-05	0.025	9	q22.2			✓	✓	✓			✓	✓			
201147_s_at	TIMP3	TIMP metallopeptidase inhibitor 3	21.5	7.E-05	0.038	22	q12.3			✓									
218711_s_at	SDPR	serum deprivation response	18.4	1.E-04	0.056	2	q32.3												
204939_s_at	PLN	phospholamban	21.9	2.E-04	0.087	6	q22.31	✓		✓	✓	✓	✓						

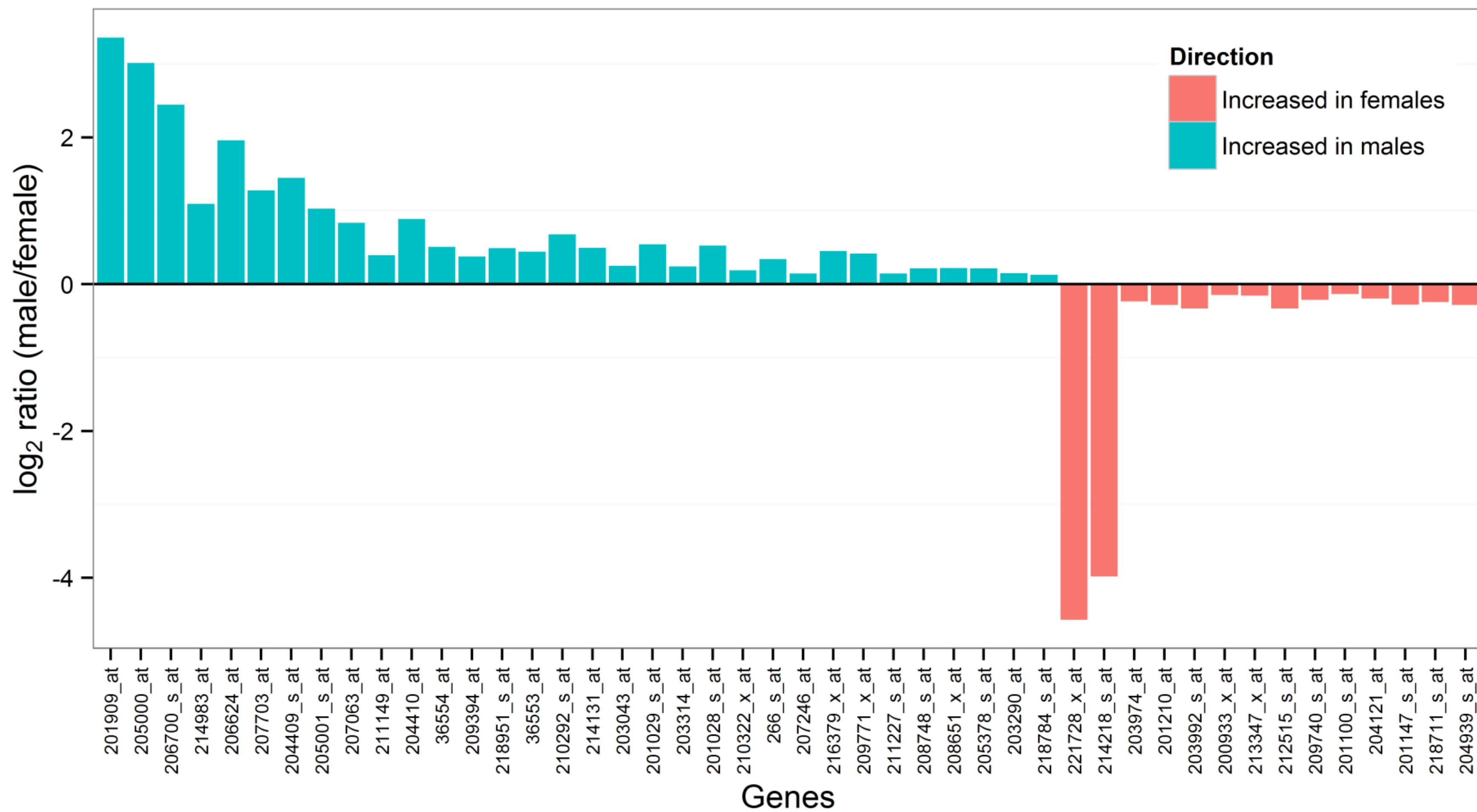


Figure 6.4. Probe sets with significantly different intensity between male and female control PFC. Probes are listed in the same order as in **Table 6.6**. Probe set intensities in blue were higher in males and probe set intensities in pink were higher in females. The log_2 ratios of PFC gene expression in males compared to females were coefficients from the linear regression.

Table 6.7. Enriched GO BP terms for genes with sex-biased expression. Criteria for enrichment are listed in **Section 6.2.6**. Genes in **orange** have increased expression in males and genes in **blue** have increased expression in females.

GO BP ID	Term	Expected significant genes	Number of significant genes	Annotated genes	P-value	Significant genes
GO:0016577	histone demethylation	0.04	3	19	9.E-06	KDM6A, UTY, KDM5D
GO:0008214	protein dealkylation	0.05	3	21	1.E-05	KDM6A, UTY, KDM5D
GO:0070988	demethylation	0.09	3	39	9.E-05	KDM6A, UTY, KDM5D
GO:0006413	translational initiation	0.30	4	135	2.E-04	DDX3X, RPS4X, RPS4Y1, EIF1AY

Table 6.8. **Probe sets with sex-schizophrenia diagnosis interactions with p<0.01.** Probe sets in orange represent male-specific differences, blue are female-specific, and white are qualitative interactions. The percentage impact on serum molecular concentrations of schizophrenia for males and females (%) from the linear regression is reported and use of robust regression is marked for each probe set. The direction of effect of potential confounders is marked with arrows. GO terms were used to assess participation of a gene in each biological process. **Abbreviations:** P (*p*-value); Q (*q*-value); A (lifetime alcohol use); AP (lifetime antipsychotic use); D (lifetime drug use); S (suicide); MP (metabolic process); DP (developmental process); CC & ST (cell communication and signal transduction); IR (defense/immune/inflammatory response); CP (cell proliferation); NSD (nervous system development). **GO terms (left to right):** GO:0008152; GO:0032502; GO:0007154 and GO:0007165; GO:0006810; GO:0040007 and GO:0008283; GO:0006950; GO:0008219; GO:0006952 and GO:0006955 and GO:0006954; GO:0007399; GO:0007268.

Affymetrix ID	Gene symbol	Gene name	Interaction (P)	Females			Males			Chromosome	Robust regression	Heterogeneity	Potential confounders	Biological processes									
				%	P	Q	%	P	Q					MP	DP	CC & ST	Transport	Growth / CP	Response to stress	Cell death	IR	NSD	Synaptic transmission
205308_at	ZC2HC1A	zinc finger, C2HC-type containing 1A	0.003	-5.5	0.23	0.29	11.3	2.E-04	0.005	8													
201063_at	RCN1	reticulocalbin 1, EF-hand calcium binding domain	0.004	-4.0	0.35	0.42	11.6	6.E-05	0.004	11					✓								
220925_at	NAA35	N(alpha)-acetyltransferase 35, NatC auxiliary subunit	0.005	-5.0	0.26	0.32	10.6	3.E-04	0.006	9								✓		✓			
202255_s_at	SIPA1L1	signal-induced proliferation-associated 1 like 1	0.006	-4.1	0.24	0.30	7.6	7.E-04	0.010	14				✓	✓	✓						✓	✓

218390_s_at	FAM204A	family with sequence similarity 204, member A	0.010	2.1	0.27	0.33	-3.6	0.002	0.016	10													
202621_at	IRF3	interferon regulatory factor 3	2.E-04	9.5	0.002	0.016	-3.9	0.026	0.065	19				✓		✓	✓		✓	✓	✓		
201219_at	CTBP2	C-terminal binding protein 2	3.E-04	10.7	0.004	0.024	-4.6	0.027	0.066	10				✓	✓			✓					
207360_s_at	NTSR1	neurotensin receptor 1 (high affinity)	7.E-04	10.2	0.002	0.016	-2.4	0.19	0.24	20						✓				✓			✓
202425_x_at	PPP3CA	protein phosphatase 3, catalytic subunit, alpha isozyme	0.002	-13.5	0.008	0.037	5.7	0.087	0.14	4			S↓	✓	✓	✓	✓		✓		✓		✓
204415_at	IFI6	interferon, alpha-inducible protein 6	0.002	-13.5	0.004	0.024	4.8	0.19	0.24	1	✓			✓		✓			✓	✓	✓		
219396_s_at	microRNA 631	MIR631	0.002	6.2	0.037	0.081	-4.4	0.010	0.041	15				✓					✓				
204844_at	ENPEP	glutamyl aminopeptidase (aminopeptidase A)	0.002	5.5	0.020	0.059	-2.9	0.042	0.084	4	✓	✓		✓	✓	✓		✓					
210231_x_at	SET	SET nuclear proto-oncogene	0.002	-7.4	0.035	0.078	5.4	0.017	0.054	3/9				✓			✓			✓			
213579_s_at	EP300	E1A binding protein p300	0.002	8.7	0.015	0.049	-3.7	0.064	0.11	22				✓	✓	✓			✓	✓	✓	✓	
216888_at	LDB3	LIM domain binding 3	0.003	7.5	0.021	0.061	-3.6	0.051	0.093	10													
204766_s_at	NUDT1	nudix (nucleoside diphosphate linked moiety X)-type motif 1	0.003	8.2	0.024	0.064	-4.2	0.041	0.084	7				✓		✓			✓				
220814_at	SLC30A6	solute carrier family 30 (zinc transporter), member 6	0.003	-4.8	0.066	0.11	4.5	0.007	0.033	2													
212475_at	AVL9	AVL9 homolog (S. cerevisiase)	0.003	5.2	0.029	0.068	-3.5	0.050	0.093	7	✓												
218020_s_at	ZFAND3	zinc finger, AN1-type domain 3	0.003	-6.4	0.041	0.084	4.6	0.022	0.062	6													
205973_at	FEZ1	fasciculation and elongation protein zeta 1 (zygin I)	0.003	10.5	0.009	0.039	-2.9	0.19	0.24	11													
205064_at	SPRR1B	small proline-rich protein 1B	0.004	6.1	0.023	0.062	-2.8	0.072	0.12	1				✓	✓								
214191_at	ICA1	islet cell autoantigen 1, 69kDa	0.004	-7.7	0.011	0.044	2.5	0.18	0.24	7													
217309_s_at	DSCR3	Down syndrome critical region gene 3	0.004	8.9	0.010	0.041	-2.5	0.20	0.25	21													
205146_x_at	APBA3	amyloid beta (A4) precursor protein-binding, family A, member 3	0.004	7.0	0.025	0.064	-3.2	0.069	0.12	19				✓	✓	✓	✓						✓
212182_at	LOC100996752	uncharacterized LOC100996752	0.004	8.8	0.13	0.18	-9.7	0.003	0.021	12			AP↑	✓		✓	✓						
217712_at	NA	NA	0.005	4.2	0.021	0.061	-1.9	0.11	0.16		✓	✓	AP↓										
217218_at	WAPAL	wings apart-like homolog (Drosophila)	0.005	6.8	0.013	0.046	-2.0	0.19	0.24	10													
206609_at	MAGEC1	melanoma antigen family C, 1	0.005	4.6	0.042	0.084	-2.6	0.046	0.089	X													
211646_at	IGH	immunoglobulin heavy locus	0.006	4.5	0.016	0.051	-1.4	0.18	0.24	14													
217442_at	CADM3-AS1	CADM3 antisense RNA 1	0.006	-9.0	0.019	0.057	3.5	0.15	0.21	1													
207670_at	KRT85	keratin 85	0.006	-5.4	0.064	0.11	4.1	0.026	0.065	12					✓								
222079_at	ERG	v-ets avian erythroblastosis virus E26 oncogene homolog	0.006	4.7	0.016	0.051	-1.5	0.19	0.24	21				✓	✓	✓		✓					
207511_s_at	CNPPD1	cyclin Pas1/PHO80 domain containing 1	0.006	5.7	0.070	0.12	-4.1	0.023	0.062	2				✓									
218372_at	MED9	mediator complex subunit 9	0.006	-5.9	0.041	0.084	3.4	0.062	0.11	17				✓									
221745_at	DCAF7	DDB1 and CUL4 associated factor 7	0.007	-8.2	0.025	0.064	3.8	0.12	0.17	17	✓		AP↓	✓	✓								
209350_s_at	GPS2	G protein pathway suppressor 2	0.007	4.7	0.075	0.12	-3.5	0.022	0.062	17				✓		✓			✓				
210683_at	NRTN	neurturin	0.007	-4.2	0.026	0.065	1.8	0.13	0.18	19				✓	✓	✓						✓	
219065_s_at	DPY30	dpy-30 homolog (C. elegans)	0.007	8.6	0.048	0.092	-4.7	0.055	0.10	2/6/													

										21													
216148_at	NA	NA	0.007	5.4	0.034	0.077	-2.4	0.095	0.15														
208748_s_at	FLOT1	flotillin 1	0.007	8.3	0.10	0.15	-7.0	0.013	0.046	6					✓	✓						✓	
214131_at	TXLNGY	taxilin gamma pseudogene, Y-linked	0.007	-10.4	0.041	0.084	6.0	0.074	0.12	Y		✓											
207655_s_at	BLNK	B-cell linker	0.008	10.6	0.075	0.12	-7.2	0.028	0.067	10			S↑		✓	✓			✓		✓		
212836_at	POLD3	polymerase (DNA-directed), delta 3, accessory subunit	0.008	5.0	0.070	0.12	-3.4	0.032	0.073	11				✓					✓				
216170_at	NA	NA	0.008	-4.5	0.070	0.12	3.3	0.032	0.073														
215521_at	PHC3	polyhomeotic homolog 3 (Drosophila)	0.008	6.4	0.023	0.062	-2.2	0.18	0.24	3					✓								
221183_at	LOC100507388	uncharacterized LOC100507388	0.008	7.1	0.044	0.086	-3.6	0.074	0.12	4													
216671_x_at	MUC8	mucin 8	0.008	6.5	0.043	0.085	-3.3	0.076	0.12														
201471_s_at	SQSTM1	sequestosome 1	0.008	5.8	0.036	0.079	-2.6	0.10	0.15	5				✓	✓	✓	✓		✓	✓			
210739_x_at	SLC4A4	solute carrier family 4 (sodium bicarbonate cotransporter), member 4	0.008	-6.5	0.11	0.16	6.5	0.013	0.046	4							✓						
218997_at	POLR1E	polymerase (RNA) I polypeptide E, 53kDa	0.008	-7.1	0.028	0.067	3.0	0.15	0.21	9				✓									
204640_s_at	SPOP	speckle-type POZ protein	0.008	-6.4	0.040	0.084	3.4	0.090	0.14	17				✓		✓			✓	✓			
201802_at	SLC29A1	solute carrier family 29 (equilibrative nucleoside transporter), member 1	0.008	6.1	0.13	0.18	-5.8	0.011	0.044	6				✓	✓	✓	✓		✓				✓
203521_s_at	ZNF318	zinc finger protein 318	0.008	5.1	0.16	0.22	-5.6	0.006	0.031	6			D↓	✓									
202400_s_at	SRF	serum response factor (c-fos serum response element-binding transcription factor)	0.008	4.4	0.11	0.16	-3.9	0.015	0.049	6				✓	✓	✓		✓	✓			✓	✓
207568_at	CHRNA6	cholinergic receptor, nicotinic, alpha 6 (neuronal)	0.009	6.8	0.051	0.093	-3.6	0.066	0.11	8						✓	✓						✓
200630_x_at	SET	SET nuclear proto-oncogene	0.009	-6.8	0.042	0.084	3.6	0.090	0.14	3/9				✓			✓			✓			
210575_at	NUDC	nudC nuclear distribution protein	0.009	4.9	0.027	0.066	-1.8	0.16	0.22														
202725_at	POLR2A	polymerase (RNA) II (DNA directed) polypeptide A, 220kDa	0.009	6.4	0.066	0.11	-4.0	0.045	0.088	17				✓					✓				
211526_s_at	RTEL1	regulator of telomere elongation helicase 1	0.009	3.1	0.071	0.12	-2.3	0.050	0.093	20	✓			✓					✓	✓			
217280_x_at	GABRA5	gamma-aminobutyric acid (GABA) A receptor, alpha 5	0.009	13.0	0.041	0.084	-7.5	0.11	0.16	15	✓	✓			✓	✓	✓		✓	✓	✓	✓	✓
220462_at	CSRNP3	cysteine-serine-rich nuclear protein 3	0.009	-10.3	0.12	0.17	11.2	0.014	0.048	2				✓						✓			
221549_at	GRWD1	glutamate-rich WD repeat containing 1	0.009	6.1	0.025	0.064	-2.0	0.19	0.24	19													
221552_at	ABHD6	abhydrolase domain containing 6	0.009	5.4	0.17	0.23	-6.2	0.006	0.031	3			S↑ AP↓	✓		✓							✓
217173_s_at	LDLR	low density lipoprotein receptor	0.010	-6.8	0.031	0.072	2.8	0.16	0.22	19				✓		✓	✓						
214164_x_at	CA12	carbonic anhydrase XII	0.010	-7.4	0.12	0.17	7.2	0.018	0.056	15				✓			✓						
215283_at	LINC00667	long intergenic non-protein coding RNA 667	0.010	6.6	0.050	0.093	-3.3	0.083	0.13	18													
213418_at	HSPA6	heat shock 70kDa protein 6 (HSP70B')	2.E-05	25.8	3.E-07	6.E-05	0.8	0.79	0.81	1	✓	✓		✓	✓				✓				

207038_at	SLC16A6	solute carrier family 16, member 6	4.E-04	-13.9	4.E-04	0.007	2.5	0.33	0.40	17							✓						
220207_at	YIF1B	Yip1 interacting factor homolog B (S. cerevisiae)	6.E-04	-11.5	2.E-04	0.005	0.8	0.69	0.71	19		✓											
213658_at	NA	NA	8.E-04	-18.7	8.E-04	0.011	3.4	0.36	0.43	15			S↓										
203802_x_at	NSUN5	NOP2/Sun domain family, member 5	9.E-04	10.7	6.E-04	0.010	-1.2	0.49	0.54	7				✓									
212284_x_at	TPT1	tumor protein, translationally-controlled 1	0.001	5.6	0.001	0.012	-0.9	0.37	0.43	13					✓	✓	✓		✓	✓			
214450_at	CTSW	cathepsin W	0.002	-12.6	1.E-04	0.004	-0.7	0.75	0.77	11				✓							✓		
221135_s_at	ASTE1	asteroid homolog 1 (Drosophila)	0.002	-10.5	0.001	0.012	1.0	0.63	0.66	3				✓					✓				
218807_at	VAV3	vav 3 guanine nucleotide exchange factor	0.002	-13.6	3.E-04	0.006	-0.4	0.85	0.86	1			S↓	✓	✓	✓	✓	✓	✓	✓	✓		
210203_at	CNOT4	CCR4-NOT transcription complex, subunit 4	0.002	10.7	1.E-04	0.004	0.9	0.58	0.62	7				✓									
216232_s_at	GCN1L1	GCN1 general control of amino-acid synthesis 1-like 1 (yeast)	0.002	10.4	0.002	0.016	-1.6	0.41	0.46	12				✓									
206287_s_at	ITIH4	inter-alpha-trypsin inhibitor heavy chain family, member 4	0.002	-8.7	0.004	0.024	2.0	0.30	0.36	3		✓		✓					✓		✓		
205599_at	TRAF1	TNF receptor-associated factor 1	0.002	9.8	0.002	0.016	-1.4	0.42	0.47	9				✓		✓				✓			
204506_at	PPP3R1	protein phosphatase 3, regulatory subunit B, alpha	0.003	-19.7	0.003	0.021	4.1	0.37	0.43	2						✓	✓		✓	✓	✓		
201854_s_at	ATMIN	ATM interactor	0.003	-10.9	0.003	0.021	2.0	0.40	0.46	16				✓					✓				
221722_x_at	LZTS1	leucine zipper, putative tumor suppressor 1	0.003	-8.0	0.003	0.021	1.2	0.46	0.52	8				✓	✓	✓			✓		✓	✓	
220858_at	SORBS2	sorbin and SH3 domain containing 2	0.003	16.1	1.E-04	0.004	1.4	0.57	0.61	4	✓		D↓										
210370_s_at	LY9	lymphocyte antigen 9	0.004	10.1	0.007	0.033	-2.3	0.28	0.34	1													
211287_x_at	CSF2RA	colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)	0.004	8.9	0.006	0.031	-1.6	0.37	0.43	X						✓							
203019_x_at	SSX2IP	synovial sarcoma, X breakpoint 2 interacting protein	0.005	-22.8	0.002	0.016	1.7	0.74	0.76	1						✓							
217177_s_at	PTPRB	protein tyrosine phosphatase, receptor type, B	0.005	6.5	0.005	0.027	-1.0	0.47	0.52	12			S↑	✓	✓								
212826_s_at	SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6	0.006	-9.3	1.E-04	0.004	-1.4	0.41	0.46	X	✓	✓	AP↑	✓		✓	✓		✓	✓	✓		
206842_at	KCND1	potassium voltage-gated channel, Shal-related subfamily, member 1	0.006	9.7	0.005	0.027	-1.1	0.57	0.61	X			AP↑			✓	✓						✓
219358_s_at	ADAP2	ArfGAP with dual PH domains 2	0.007	12.5	0.004	0.024	-1.1	0.66	0.69	17				✓	✓								
220652_at	KIF24	kinesin family member 24	0.007	-6.4	0.005	0.027	0.7	0.60	0.64	9													
204277_s_at	TK2	thymidine kinase 2, mitochondrial	0.007	-6.8	0.013	0.046	1.9	0.27	0.33	16				✓									
214638_s_at	CCNT2	cyclin T2	0.007	8.0	7.E-04	0.010	0.6	0.65	0.68	2				✓		✓							
209105_at	NCOA1	nuclear receptor coactivator 1	0.008	15.1	0.007	0.033	-2.1	0.49	0.54	2				✓	✓	✓	✓			✓		✓	
205600_x_at	HOXB5	homeobox B5	0.008	8.2	0.012	0.046	-1.8	0.34	0.41	17				✓	✓								
204939_s_at	PLN	phospholamban	0.008	-17.5	0.002	0.016	-0.3	0.94	0.95	6	✓	✓		✓	✓	✓	✓						

201819_at	SCARB1	scavenger receptor class B, member 1	0.008	10.7	0.005	0.027	-0.9	0.66	0.69	12				✓		✓	✓	✓	✓			
214327_x_at	TPT1	tumor protein, translationally-controlled 1	0.008	5.5	0.002	0.016	0.0	0.97	0.97	13			S↑		✓	✓	✓		✓	✓		
217970_s_at	CNOT6	CCR4-NOT transcription complex, subunit 6	0.009	7.4	0.019	0.057	-2.1	0.24	0.30	5			D↓	✓				✓				
216475_at	TUBB4BP7	tubulin, beta 4B class IVb pseudogene 7	0.009	-8.0	0.014	0.048	1.9	0.35	0.42	6												
208885_at	LCP1	lymphocyte cytosolic protein 1 (L-plastin)	0.009	14.4	0.003	0.021	0.0	0.99	0.99	13					✓	✓	✓				✓	
211577_s_at	IGF1	insulin-like growth factor 1 (somatomedin C)	0.009	8.8	0.013	0.046	-1.8	0.38	0.44	12				✓	✓	✓	✓	✓	✓	✓		✓
217349_s_at	PRICKLE3	prickle homolog 3 (Drosophila)	0.010	8.5	4.E-04	0.007	1.0	0.53	0.58	X	✓	✓										
203662_s_at	TMOD1	tropomodulin 1	0.010	-11.2	0.010	0.041	1.9	0.49	0.54	9		✓	S↓		✓							
213842_x_at	NSUN5P1	NOP2/Sun domain family, member 5 pseudogene 1	0.010	7.0	0.013	0.046	-1.4	0.39	0.45	7				✓								
205396_at	SMAD3	SMAD family member 3	0.010	10.0	0.009	0.039	-1.4	0.52	0.57	15				✓	✓	✓	✓	✓	✓	✓	✓	
218737_at	SBNO1	strawberry notch homolog 1 (Drosophila)	0.010	-9.6	0.009	0.039	1.0	0.59	0.63	12	✓			✓								

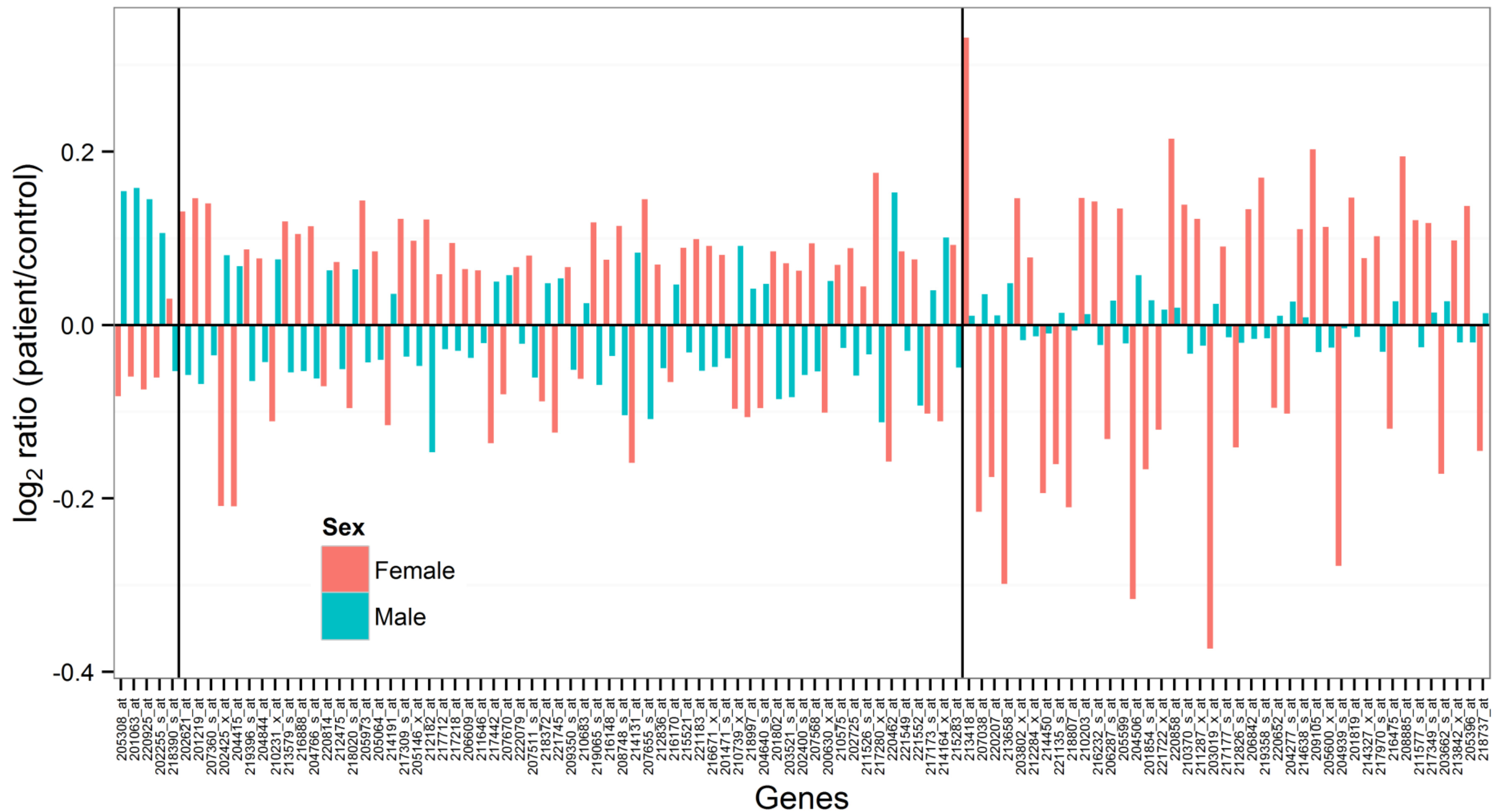


Figure 6.5. Probe sets with sex-schizophrenia diagnosis interactions with *p*<0.01. Probes are listed in the same order as in **Table 6.8**. Black vertical lines separate male-specific differences, female-specific differences, and qualitative interactions. The $\log_2$ ratios of PFC gene expression in patients compared to controls for males (blue) and females (pink) were coefficients from the linear regression.

6.4 Discussion

6.4.1 Differential PFC gene expression in schizophrenia

The current work takes a novel approach to genome-wide expression analysis in the PFC by comprehensively investigating the effects of sex, diagnosis, and sex-diagnosis interactions in a meta-analysis of schizophrenia patients and control subjects. The agreement between these results and serum protein concentration in first onset, antipsychotic naive schizophrenia patients and controls is also examined. Previous studies have investigated differential expression of genes in schizophrenia or sex separately, and have not considered interactions.

This study used a different approach from either *Mistry, Gillis, and Pavlidis (2012)* (*430*) or *Pérez-Santiago et al (2012)* (*429*) to meta-analyse schizophrenia and control PFC gene expression data. Although only a subset of the cohorts from the two previous meta-analyses was used, 209 differentially expressed genes (229 probe sets) were detected, compared to 95 differentially expressed genes (125 probe sets) in *Mistry, Gillis, and Pavlidis (2012)* (*430*) and 144 (160 probe sets) in *Pérez-Santiago et al (2012)* (*429*). In particular, the use of array weights and robust regression in the presence of regression outliers in this study increased the number of differentially expressed genes detected by over 13 fold. Similar to both of the previous studies, more downregulated than upregulated probe sets were found. The relatively small intersection of findings across all meta-analyses (13 probe sets) suggests that different datasets used, pre-processing methods, detection and handling of poor quality arrays, modelling, and FDR calculations between the analyses have a considerable impact on results. High heterogeneity of the additional cohorts included in the previous meta-analyses may contribute to the discrepancies and overall smaller number of differentially expressed genes found previously. However, after sequentially removing each dataset and re-analysing the remaining data, *Mistry, Gillis, and Pavlidis (2012)* (*430*) reported that their results were robust and not highly influenced by any dataset. They also found a good agreement between their results and results found using different methods for pre-processing, including variations on RMA and MAS5 algorithms (*430*). Sources of discrepancy may therefore arise from other dissimilarities between analyses. Nevertheless, a considerable number of the current findings (30% of probe sets) overlapped with findings from at least one of the previous meta-analyses. Included in these findings were BAG3, C4B, MT1X, NEUROD6, and EGR1, which were confirmed in an independent cohort by *Pérez-Santiago et al (2012)* (*429*) to be differentially expressed in schizophrenia patients. Other findings were consistent with reports from individual cohorts, were a different probe set from a reportedly differentially expressed gene, or have been replicated independently or

in separate, more specific studies of SMRI or HBTRC brains. For example, NR3C1 (glucocorticoid receptor) expression was decreased in schizophrenia patients in other studies of SMRI samples and in several brain regions, including the dorsolateral prefrontal cortex (DLPFC), hippocampus, inferior temporal cortex, and amygdala, and in bipolar disorder and depression (*444*, *445*). Reduced BDNF gene expression found in this meta-analysis confirms previous findings in the DLPFC in an independent cohort (*446*) and in blood levels of the protein (*210*).

Several new findings in this meta-analysis yielded novel insight into processes disturbed in schizophrenia and provide more evidence for previous lines of inquiry. OXTR gene expression was higher in schizophrenia in the present meta-analysis, but was not reported as being differentially expressed in any of the individual cohorts or previous meta-analyses. Oxytocin acts as a hormone and neurotransmitter, regulates social behaviours and cognition, and interacts with gonadal hormones, dopamine, serotonin, and glutamate (*447*, *448*). Recent studies have found polymorphisms of the OXTR gene to be associated with risk for schizophrenia (*449*), higher plasma oxytocin levels to be negatively associated with general and positive symptoms of schizophrenia in females (*450*), and adjunctive intranasal oxytocin therapy in patients to improve Positive and Negative Syndrome Scale (PANSS) scores (*451*). GHR gene expression was also increased in schizophrenia. The growth hormone (GH)/insulin-like growth factor (IGF)-1 axis is important in regulating brain function, including cognitive functioning (*452*). Decreased GH has been found in both the pituitary of chronic schizophrenia and serum of first onset schizophrenia patients (*209*, *359*). These findings provide further support for involvement of GH and oxytocin in schizophrenia. Other findings included increased CADM1, which can induce synapse formation (*453*), and reduced SYNDIG1, which regulates the content of α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors at synapses and may be involved in synaptic plasticity (*454*).

A GO term enrichment analysis revealed novel terms associated with schizophrenia and consistent with the literature. Disturbed cell signalling has been observed in schizophrenia and this study provided further support for this by finding the GO term ‘cellular response to cAMP’ enriched. Abnormal activity of cAMP-associated signalling pathways has been reported in the DLPFC in schizophrenia, implicating a disturbance in the ability to combine information from neurotransmitter receptors and, together with abnormalities in MAPK-associated pathways, extensive signalling deficits in schizophrenia (*455*). GO terms involving response to metal ion/transition metal nanoparticle and neuronal apoptosis are consistent with a previous pathway-based analysis of genome-wide association study (GWAS) data in schizophrenia (*456*) and with the results of a GO enrichment analysis

performed by *Narayan et al (2008; GSE21138 cohort)* (*198*) that found enriched terms related to metal ion binding and apoptosis. Additional evidence was found here that processes controlling the release of cytochrome c from the mitochondria may be involved in neuronal apoptosis in schizophrenia (*457–460*). Indication of mitochondrial dysfunction was further found by enrichment of the 'respiratory electron transport chain' GO term, which supports a parallel transcriptomic, metabolomic, and proteomic study of SMRI brains linking disturbed metabolism and oxidative stress to schizophrenia (*51*). Novel detection of differentially expressed genes belonging to these GO terms may provide greater insight into pathophysiological mechanisms related to these functions in schizophrenia. Also enriched were 'response to steroid hormone stimulus' and 'response to estradiol' terms. Consistent with the *Perlman et al (2005)* (*461*) study of estrogen receptor (ER) α mRNA in the DLPFC, however, there were no significant differences or sex-dependent differences in ERα or ERβ gene expression in schizophrenia patients in the present work. However, pathway analyses of GWAS data have previously found the 'CARM1 and regulation of the estrogen receptor' (BioCarta) pathway (*462*), 'androgen and estrogen metabolism' (KEGG) pathway (*462*), and the 'estrogen biosynthesis process' (GO) term overrepresented (*463*). Furthermore, *Weickert et al (2008)* (*464*) found an ERα genotype associated with schizophrenia risk and lower mRNA levels in the frontal cortex, more novel ERα SNPs in schizophrenia, and more ERα splice variants expressed in schizophrenia, one of which reduced gene expression at estrogen-response elements. Further work into the role of estrogen in schizophrenia in the PFC is warranted to explain the current results in the context of these previous findings. Finally, though not specifically enriched, a number of differentially expressed genes were also related to metabolic and developmental processes; cell communication and signal transduction; transport; growth and cell proliferation; stress response; cell death; defence, immune, and inflammatory response; and nervous system development and synaptic transmission, which have previously been found to be relevant to schizophrenia.

Potential confounding effects and heterogeneity between cohorts should be considered when interpreting these results. This study showed that differential expression of a number of genes including S100A8, C4B, GHR, and others may be caused by exposure to antipsychotic medication, lifetime drug or alcohol use, or suicide in patients. *Pérez-Santiago et al (2012)* (*429*) also found MT1X increased and others (27 of 144 significant genes) differentially expressed in fetal brain aggregates treated with haloperidol or clozapine. However, their results were based on pooled samples measured on only one Affymetrix HG-U133 Plus 2.0 chip for each group. S100A8 and S100A9 have also been found to be increased in rat frontal cortex after chronic treatment with olanzapine (*465*). A small fraction of probes (14%) showed evidence for heterogeneity across cohorts, including S100A8 and

S100A9, C4B, and CADM1. Sources of heterogeneity should be investigated and are discussed further in the sections below.

Finally, seven serum protein with altered concentrations in first onset, antipsychotic naive schizophrenia patients also showed differential gene expression in PFC tissue in the same direction (p<0.05). This was a surprising finding, given that this post-mortem PFC meta-analysis was conducted in different subjects, at different stages of illness, with different exposures to antipsychotic medication and other confounding factors, and it measured gene expression. These findings would need to be validated as some had high q-values and only the most significant probe set from a gene was reported. The serum proteins had functions in immune/inflammatory response (ferritin, CEA, EN-RAGE, α1-antitrypsin) and apoptosis (CA 19-9, HGF, TIMP-1). These findings may suggest that peripheral tissue can reflect certain elements of brain dysfunction in schizophrenia. Previous studies have also found parallel brain and serum alterations in gene expression and/or protein and small molecule concentrations in schizophrenia (*213*, *214*). *Harris et al (2012)* (*213*) has proposed that this is due to coordination of the brain and periphery through hormones and other regulatory molecules.

6.4.2 Sex-biased gene expression

Sex-biased genes were largely located on the sex chromosomes. Many of the sex-biased genes found here have previously been identified, including XIST, KDM5D, RPS4X, HDHD1, ASMTL, RPS4Y1, NLGN4Y, TTTY15, PCDH11Y/PCDH11X, DDX3Y, USP9Y, UTY, ZFY, CD99, DDX3X, USP9X, and EIF1AY (*436–438*, *466*). There were no significant sex differences in expression of ERα in the PFC, in agreement with *Perlman et al (2005)* (*461*). ERβ also did not show sex-biased expression here. Several of the genes with higher expression in males were X-Y homologous genes. *Weickert et al (2009)* (*437*) found that for PCDH11Y/PCDH11X this was due to higher PCDH11Y expression in males, suggesting that there were no compensatory increases in female PCDH11X expression. Enriched GO terms in sex-biased terms included histone demethylation and transcription initiation. Previous studies have suggested that abnormal X-chromosome inactivation may occur in psychiatric disorders, with higher XIST, KD5MC, and other X-chromosome gene expression found in lymphoblastoid cells of female psychiatric patients with mania, psychosis, and major depression (*467*). However, this finding was not replicated here for PFC gene expression in schizophrenia. None of the sex-biased genes were differentially expressed in schizophrenia, but two (FLOT1 and PLN) had sex-diagnosis interactions with p<0.01. FLOT1 has functions in cytoskeletal interactions, endocytosis, and signalling (*468*), and was male-biased in control subjects with reduced expression in male schizophrenia patients compared to

controls, relative to female differences. Expression differences in PLN may be due to cross-hybridization as it is not known to be expressed in the brain (*469*). Finally, very few of the serum analytes with different concentrations between males and females (**Chapter 3**) were also differentially expressed in the PFC. This work and other works examining sex-biased gene expression in the PFC contained a limited number of subjects (121 in the current study), and here there was a heavy male bias (69%). Subjects in this post mortem study were older than in the serum study, and female hormonal status at time of death was not available. Tissue was also taken from different regions in the PFC, which potentially contributed to the heterogeneity across cohorts found in a large proportion (approximately 40%) of sex-biased genes.

6.4.3 Sex-schizophrenia diagnosis interactions in PFC gene expression

No genes had a significant sex-schizophrenia diagnosis interaction (q<0.1) in this study, despite the GO term 'response to estradiol' being enriched in differentially expressed genes in schizophrenia. The high proportion of males included in each of the cohorts (69%) meant the power to detect significant differences between female schizophrenia patients and control subjects was reduced, and that the differentially expressed genes detected here were based largely on male data. Future studies must recruit more female subjects to ensure that the findings presented here and in other publications based on these data adequately represent differential gene expression in both sexes. Many of the sex-diagnosis interactions examined (p<0.01) had female-specific differences in gene expression in schizophrenia (40% of interactions, compared to less than 5% male-specific differences) that warrant further investigation in a larger cohort.

Although there were no significant GO terms arising from genes with sex-diagnosis interactions (p<0.01), several of the genes participated in processes previously found to be important in schizophrenia. Genes involved in apoptosis included the interferon induced IFI6, which was reduced in female schizophrenia patients and plays an important role in regulating mitochondrial-mediated apoptosis by inhibiting release of cytochrome c (*470*). Furthermore, expression of PPP3R1 and PPP3CA (protein phosphatase 3, regulatory subunit B, α and protein phosphatase 3, catalytic subunit, α isozyme, respectively) were decreased in female schizophrenia. High levels of calcineurin (made up of subunits encoded by these genes) have been found to induce cytochrome c-dependent apoptosis in neurons (*471*). These may play a role in sex-dependent differences in PFC volume in schizophrenia, although these findings vary by study and PFC region (*249*, *255*, *256*). Calcineurin also plays an important role in immune response, and inhibitors of calcineurin are potent immunosuppressors (*472*). Mice with forebrain-specific calcineurin knockout exhibit impaired

working memory (*473*) and schizophrenia-like behaviours (*474*). Furthermore, NTSR1 expression was increased in the PFC of female schizophrenia patients. This is in contrast with NTSR1 knockout mice that exhibited similar behavioural and frontal cortex molecular abnormalities to the chronic phencyclidine (PCP)-treated mouse model of schizophrenia (*475*). Neurotensin is hypothesized to be an endogenous antipsychotic compound implicated in the pathophysiology of schizophrenia (*476*), which interacts with dopamine systems and is regulated by estrogen (*477*). Important sex differences in the role of neurotensin and calcineurin in the pathophysiology of schizophrenia may exist.

None of the proteins with significant sex-diagnosis interactions in serum concentration were found to have parallel sex-diagnosis interactions in PFC gene expression ($p<0.05$). In addition to limitations described above, sex and gender differences in the course of illness, age of onset, and effects of antipsychotic medication (*243*) may further contribute to discrepancies in gene expression findings. However, sex-dependent differences in gene expression related to androgen receptor signalling and defence/immune response were found in the PFC tissue of schizophrenia patients. This was of particular relevance given the increased serum testosterone found in first onset female schizophrenia patients and increased immune response markers in male schizophrenia patients found in **Chapter 5**. EP300 and NCOA1 expression was increased in the PFC of female schizophrenia patients and are important for the production and operation of the androgen receptor (AR) transcription complexes (*478*). Androgens have been reported to increase EP300, NCOA1, and IGF-1 in prostate cancer (*478*, *479*) and were also over-expressed in female schizophrenia patient PFC tissue. However, androgen receptor expression itself and ERα and ERβ expression were not modified in a sex-specific manner. As discussed above, this is consistent with *Perlman et al (2005)* (*461*)*,* who also found no significant differences between male or female schizophrenia and control subjects in ERα mRNA expression in the DLPC. A number of genes, including IRF3, IFI6, PPP3R1, PPP3CA, ITH4, CTSW, and BLNK were involved in immune response and may also provide further indication of sex differences in immune dysfunction in schizophrenia. However, in contrast to serum results, many of these were female-specific.

A number of limitations must be considered when interpreting results from this study. Gene expression was measured in these microarray studies, which may not correlate with protein concentration. In addition, a number of different brain areas were used in this meta-analysis, including BA9, 10, and 46, which may contribute to heterogeneity in 14%, 39%, and 10% of the results for analyses of schizophrenia diagnosis, sex, and sex-diagnosis interaction effects, respectively. Different diagnostic instruments/protocols, physicians, inclusion/exclusion criteria, experimental procedures, age of subjects and course/length of

illness, and confounding factors may also create heterogeneity between cohorts. Lifetime substance use, antipsychotic use, and suicide status are confounding variables that can influence gene expression and for which little power was available to detect effects. Other demographic, health, and lifestyle differences may further confound differences in gene expression between schizophrenia and controls, but were unavailable or unrecorded. Findings should therefore be validated in a better characterized cohort. As previously mentioned, hormonal status was not recorded for female subjects and other differences between males and females in course of illness, age of onset, and effects of antipsychotic medication (*243*) may be confounded with sex.

In summary, the current study used a meta-analysis of five cohorts of microarray data to explore the effects of schizophrenia diagnosis, sex, sex-schizophrenia diagnosis interactions, and confounding effects of suicide, and lifetime drug, alcohol, and antipsychotic medication use on PFC gene expression. The use of array weights and robust regression resulted in a greater capacity to detect differentially expressed genes in schizophrenia, confirming a number of prior findings and resulting in several novel ones. Findings included genes enriched in processes related to neuronal apoptosis, mitochondrial function, and response to estradiol and cAMP. Some results may also mirror differences in serum protein concentrations between first onset, antipsychotic naive schizophrenia patients and controls. Finally, although no sex-diagnosis interactions in gene expression reached significance ($q<0.10$), genes with $p<0.01$ were involved in mitochondrial-mediated apoptosis and synaptic transmission important in schizophrenia and in androgen receptor signalling and immune response, which may have sex-dependent roles in schizophrenia. However, as these had high q-values, further validation of these results is necessary, together with elucidation of their potential role in sex-specific pathophysiological mechanisms and in clinical and epidemiological sex and gender differences in schizophrenia. Importantly, more females should be included in microarray studies of the PFC and other brain regions in order to have more power to detect sex and sex-diagnosis interaction effects in gene expression.

Chapter 7 Final discussion

This thesis used serum and brain molecular profiling to investigate sex and female hormonal status in controls, schizophrenia, and major depressive disorder (MDD). The aim of this chapter is to provide an overview of the major findings resulting from this work, limitations of the studies, and implications for future investigations.

7.1 Summary of major findings

7.1.1 Inflammation and immune response

Many of the sex-dependent serum biomarkers of MDD found in **Chapter 4** were associated with increased inflammation, immune, and/or defence response in males. These included C-reactive protein (CRP), trefoil factor 3 (TFF3), cystatin-C, fetuin-A, β2-microglobulin, CD5 antigen-like (CD5L), and tumor necrosis factor receptor 2 (TNFR2). Similarly, elevated serum levels of interleukin (IL)-15, α1-antitrypsin (AAT), IL-8, macrophage derived chemokine (MDC), ferritin, intercellular adhesion molecule 1 (ICAM-1), and macrophage inflammatory protein-1 α (MIP-1α) involved in inflammation and immune response were found in first onset, antipsychotic naive male schizophrenia patients only in **Chapter 5**. These were accompanied by findings of potential sex-dependent differences in expression of genes involved in immune response in the prefrontal cortex (PFC) in schizophrenia patients in **Chapter 6**. Extensive differences in the serum levels of proteins involved in these biological processes were observed between males and females using oral contraceptives (OCs), after menopause, and with menstrual cycles in **Chapter 3.** These included higher CRP, TFF3, AAT, and fetuin-A in the serum of female OC users compared to females with menstrual cycles and higher cystatin-C, MDC, and ferritin in males compared to females. Brain and peripheral effects of estrogen, progesterone, and testosterone on immune and inflammatory response have also been reported elsewhere (see **Chapter 1**).

These results require further validation and investigation, but may have implications for the current hypotheses of depression and schizophrenia. Immune hypotheses of schizophrenia and depression could have a male-specific component that may contribute to sex and gender differences in incidence, prevalence, age of onset, and other clinical characteristics. Inflammation may contribute to hypothalamic-pituitary-adrenal (HPA) axis dysregulation and can affect neurotransmission (*143*). Pro-inflammatory cytokines can elicit apoptotic activity (*480*) and expression of genes involved in apoptosis was disturbed in the PFC in schizophrenia

(**Chapter 6**). Previous work has already suggested that higher CRP occurs specifically in male depression (*124*, *167–171*) and that there are sex differences in the associations between CRP and fibrinogen and burnout, depression, and anxiety (*169*). Other lines of evidence also point to potential male-specific disruptions in inflammation in mental disorders. Asperger syndrome in adults has been characterized by increased serum concentration of some immune and inflammatory molecules in males only (*10*). A mouse model of early prenatal stress, a risk factor for schizophrenia and autism spectrum disorders, causes male-specific alterations in dopamine D1 and D2 receptor gene expression, together with indications of placental inflammation and stress-induced locomotor hyperactivity that were partly resolved by maternal treatment with non-steroidal anti-inflammatory drugs (NSAIDs) (*481*). Prenatal exposure to valproic acid used to produce a rat model of autism has also shown more male behavioural and immunological abnormalities (*482*) and stressor re-exposure in mice produces male-specific sensitized expression of IL-1β mRNA in both the PFC and hippocampus (*483*). Sex differences in immune regulation (see also **Chapter 1**) may be important in understanding these findings. These are widespread and were shown in **Chapter 3** to involve many of the analytes with sex-dependent differences in schizophrenia and MDD. Males may be more prone to dysregulation of acute inflammation and proinflammatory immune response, as suggested by the higher prevalence of males with autoimmune diseases with these characteristics (*380*). Androgen activation and estrogen suppression of T-lymphocytes have been proposed to account for some of the sex differences in schizophrenia in the macrophage-T-lymphocyte theory of schizophrenia (*225*). Sex differences have also been found in microglial cell morphology, function, and number (see **Chapter 1**), including anti-inflammatory effects of estrogen on their activation (*55*, *56*). Three-way interactions between the HPA and hypothalamic-pituitary-gonadal (HPG) axes and the immune system may also contribute to sex-dependent differences in immune and inflammatory markers in schizophrenia and MDD (*29*, *355*, *376*, *377*). Further work is needed to elucidate mechanisms for the potential sex differences in immune and inflammatory dysregulation in mental disorders.

Sex differences in response to treatment with anti-inflammatory medication may also be expected if these results reflect causal mechanisms of schizophrenia and MDD. However, as discussed in **Chapter 5**, a meta-analysis of 26 trials on anti-inflammatory treatment efficacy in schizophrenia did not show results stratified by sex (*423*), although a lower risk of psychosis has also been found specifically in male NSAID (*421*) and glucocorticoid (*422*) users. Another meta-analysis of 14 trials of anti-inflammatory treatment in depression reported a reduction in depressive symptoms compared to placebo, but also did not report results by sex (*484*).

7.1.2 HPG axis

The HPG axis affects development and function of the nervous system and interacts with other components of the endocrine system (see **Chapter 1**). In this work, alterations in the HPG axis were found in schizophrenia and MDD. In schizophrenia, increased FSH, LH, and progesterone were found together with female-specific reduction of SHBG and elevation of testosterone in serum (**Chapter 5**). Differentially expressed genes in the PFC of schizophrenia patients were enriched in functions related to cellular response to steroid hormone stimulus and response to estradiol (**Chapter 6**) and expression of EP300 and NCOA1 genes, important for production and function of androgen receptor transcription complexes, was increased specifically in PFC tissue in female schizophrenia patients (**Chapter 6**). In MDD, comorbid MDD and anxiety disorder (s) (CMA), and remitted MDD, elevated testosterone and/or reduced FSH were found in females (**Chapter 4**). These observations were consistent with previous findings indicating disturbed sex dimorphisms in schizophrenia (*277*) and provide support for hypotheses that reproductive hormones may be involved in MDD (*164*).

Periods associated with fluctuating hormonal levels affect risk for female MDD and schizophrenia. While estrogen treatment and periods of high estrogen have been associated with improved symptoms of schizophrenia (*9*, *260–263*), periods of low estrogen have been linked to psychosis (*260*). Affective disturbances in females may coincide with rising hormone levels during puberty or be cycle-dependent, and perimenopause and postpartum are periods of increased risk for depression (*156*, *160*). Together with genetic and environmental factors during development, fluctuating hormonal levels that occur throughout female life associated with puberty, the menstrual cycle, OC use, menopause, and pregnancy may create susceptibility for hormonal dysregulation and vulnerability to mental disorders. The role of testosterone in female schizophrenia and depression has been less explored compared to other reproductive hormones, but has been shown in this work to be of potential importance to these disorders (further discussed in **Chapters 4-5**). Excess androgens and menstrual disturbance are characteristics of polycystic ovary syndrome (PCOS), the most common endocrine disorder in reproductive-age women (affects 4-8% of women) (*485*). A higher prevalence of depression and autism spectrum (*486*), anxiety, bipolar, and eating disorders has been found in females diagnosed with PCOS (*485*). Several common characteristics of PCOS also occur in mental disorders, including insulin resistance, increased HPA axis activity, and elevated LH (*392–395*). Although the aetiology of PCOS has not been elucidated, genetic and environmental factors have been implicated that may disrupt fetal organization through excess prenatal androgen, leading to development of the PCOS phenotype in adulthood (**Figure 7.1**). Disrupted fetal hormonal programming has been linked to MDD (*164*), schizophrenia (*277*), and autism spectrum disorders (*487*), which are also characterized by higher serum testosterone in adult

females with Asperger syndrome (*10*). These disorders may share elements of their aetiologies and pathophysiologies with PCOS. Further work should explore these links and potential treatment options.

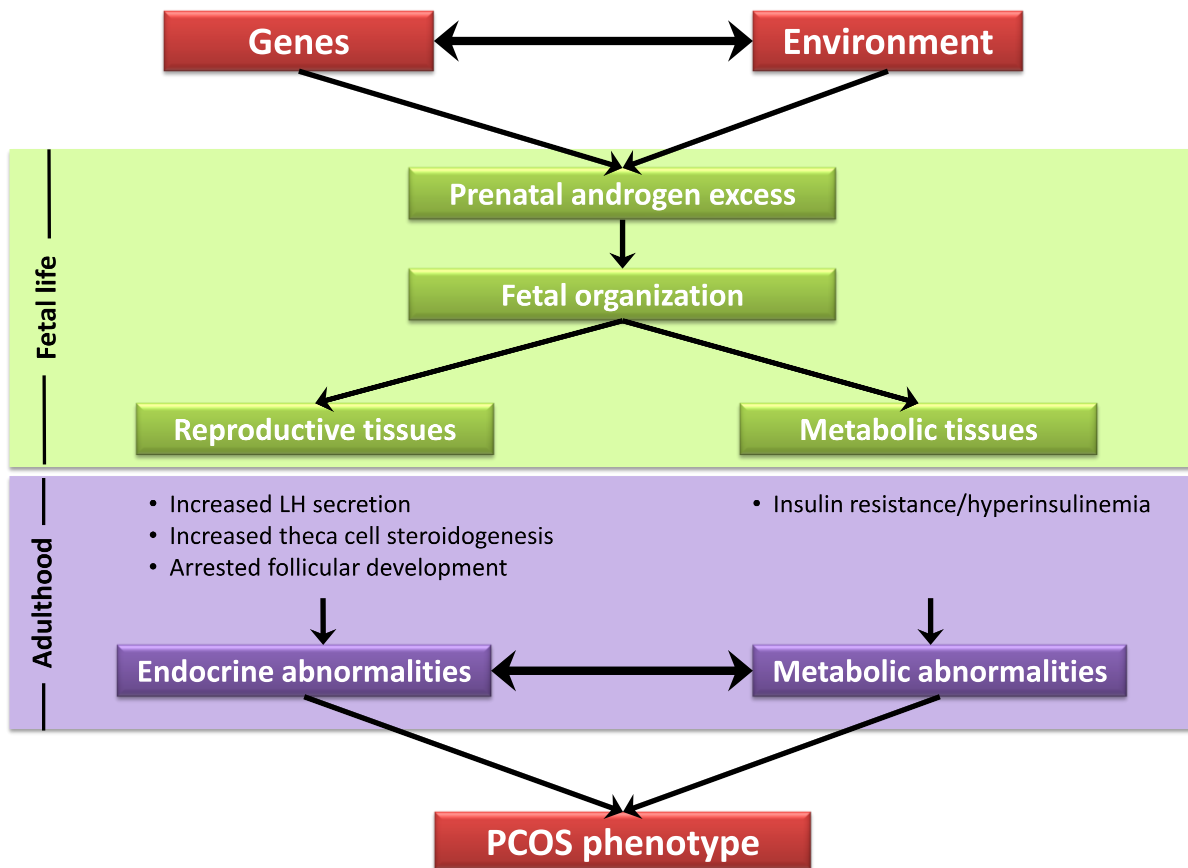


Figure 7.1. Model of development of the PCOS phenotype. Genetic and environmental factors may influence fetal development to produce an adult PCOS phenotype. Based on *Xita and Tsatsoulis (2006)* (*488*) and *Crain et al (2008)* (*489*).

7.1.3 Implications for biomarker and microarray studies

The results of this work have several implications for future biomarker and microarray studies. In particular, it was shown that sex and female hormonal status are important factors that should be explicitly taken into account in future investigations. Individual chapters discuss these implications in more detail, but they are summarized here.

Results from **Chapter 3** showed that biomarker studies should match groups based on both sex and female hormonal status. In this work, the presence of extensive differences between males and females with menstrual cycles, using OCs, and after menopause in serum concentrations of molecules involved in endocrine, immune, metabolic, and other processes was demonstrated. These extensive differences had a large impact on the number of false positive results in groups with simulated imbalances in the proportions of males or OC users. Group imbalances may occur more often in mental disorders, where sex differences in the prevalence of disorders

occurs and OC use may be less consistent (*366*). Random forest classifiers developed in **Chapter 3** may provide a means of retrospectively investigating group imbalances in previously conducted studies.

Chapters 4 and **5** showed that sex-dependent serum molecular differences were present in comparisons between controls and first onset, antipsychotic naive schizophrenia and MDD patients. It was demonstrated that only male MDD patients could be classified against controls with an accuracy greater than chance using the investigated panel of molecules, though performance was still poor. More work is needed to develop a viable set of biomarkers for male and female MDD. A previous study also found that sex-specific serum biomarker panels classified adults with Asperger syndrome from controls (*10*). Future biomarker studies of mental disorders should consider sex and gender in their design and analysis to improve reproducibility of findings.

Chapter 6 showed that array weights and robust regression could be used to improve the capacity to detect differentially expressed genes in a microarray study of PFC tissue in schizophrenia patients, resulting in novel findings. Parallels were found between PFC gene expression and some of the serum findings in first onset, antipsychotic naive schizophrenia. However, the small number of females included in these studies made it difficult to assess sex differences and sex-diagnosis interactions in gene expression. Post-mortem PFC studies should include more females in order to ensure the extensive microarray findings in schizophrenia reported in the literature (discussed in **Chapters 1** and **6**) are relevant to both males and females.

7.2 Limitations and future outlook

A number of the limitations of this body of work are addressed in this section. In particular, these results will need to be validated in independent cohorts. A number of the findings in **Chapters 4-6** had high q-values ($q>0.25$), including eight of the sex differences in serum analyte associations with log-odds of MDD (**Chapter 4**), four of the sex-dependent differences in serum concentration in schizophrenia patients (**Chapter 5**), and all sex-dependent differences in PFC gene expression in schizophrenia patients (**Chapter 6**). As these studies were not specifically designed to investigate sex, more appropriate planning should occur before conducting further investigations of sex-dependent differences and other sources of heterogeneity in these disorders. Effort should be taken to recruit adequate numbers of both male and female patient and control subjects to serum and PFC studies and to collect and assess precise information about female hormonal status, the presence of sex chromosome

abnormalities, and diagnostic, health, lifestyle, and socioeconomic variables known to differ between genders. Other limitations associated with cohort characterization, the type and quantity of molecular data collected from participants, statistical methods, and other more fundamental problems that occur in the study of mental disorders are discussed below. Individual chapters address limitations of the work specifically covered in them.

Cohorts used in the schizophrenia serum and post-mortem PFC studies (**Chapters 5** and **6**) included limited demographic and other information. The information that was collected was heterogeneous between cohorts. It was therefore not possible to account for variability in serum concentration or gene expression associated with these factors in the analyses. **Chapter 3** showed the importance of accounting for female hormonal status and future work should further assess the impact of other important variables, including smoking, drug and alcohol use, ethnicity, use of psychiatric and other medications (short and long term), history and comorbidity with mental disorders and other diseases, body mass index, age, blood pressure, fasting status, family history, socioeconomic, and other demographic, health, and lifestyle variables on serum concentration and PFC gene expression. This may help to explain heterogeneity in results of the meta-analyses and to distinguish between effects of sex and gender. Much of the information collected in the Netherlands Study of Depression and Anxiety (NESDA) in **Chapters 2-4** was assessed by self-report, including female hormonal status and lifestyle variables. These variables may be inaccurate and in some cases may suffer from social desirability bias (*490*). Furthermore, molecular data for these studies (**Chapters 3-6**) were collected at a single time point. Circadian changes in serum molecular levels and changes associated with other events were therefore not evaluated. Disruptions of circadian clocks (*491*) and response to psychological stress (*139*) have been linked to mental disorders and these may therefore be important to examine in future studies. Changes associated with reproductive events throughout the menstrual cycle, pregnancy and the postpartum period, transition to menopause, and OC use could also not be assessed, some of which may represent a period of risk or remission for MDD and schizophrenia. Furthermore, the effects on serum molecular concentrations of changes in diagnoses, long term use of psychiatric medication, and course of illness could also not be investigated. Future studies should evaluate these in order to gain greater insight into potential pathophysiological mechanisms.

Several limitations were also associated with microarray and multiplex immunoassay data collected from participants (see also **Chapter 2**). The number of molecules investigated in these platforms makes individual inspection of each difficult or impossible. Data may consequently be modelled inappropriately, resulting in false positive results or in important results being missed. Substantial batch effects were also found in both microarray and multiplex immunoassay data and several poor quality arrays were present in microarray data (**Chapter 6**). Use of different

methods to find multivariate outliers, determine use of robust regression, transform data, account for multiple testing, and impute missing data may affect results from **Chapters 3-5** and these should be further investigated in future. The *Benjamini and Hochberg (1995)* (*326*) false discovery rate (FDR) procedure used to control the expected proportion of false positive findings in this work assumes independent tests, violated here by the correlations between molecular concentrations and expression of genes. Although the procedure has been shown to control the FDR in simulated DNA microarray data (*492*), other procedures including the *Benjamini and Yekutieli (2001)* (*493*)*, Storey and Tibshirani (2003)* (*494*)*,* and resampling-based procedures (*492*) should also be tested. These procedures have been reported to control the FDR and to improve power for the latter two. In microarray data, poor correlations between mRNA and protein levels have been reported and this limits the conclusions reached from these investigations. On the other hand, though multiplex immunoassays measure protein and other small molecule concentrations, the prohibitive costs associated with incorporating new assays into the platform limits the number of measured molecules. Estrogen was not measured in serum studies (**Chapters 3-5**), which would have provided more insight into disruptions of the HPG axis and the estrogen hypothesis of schizophrenia. The proteins measured were biased toward particular biological processes. These included regulation of cell motility, response to stress, regulation of signal transduction, immune response, and others (**Table 7.1**). Future studies should incorporate molecules involved in a broader range of biological functions and perform pathway analysis. Furthermore, post-translational modifications, alternatively splicing, and cleaved proteins cannot be distinguished in the current panel of molecules. Finally, integration of this data with other genomic, metabolomic, and neuroimaging data would provide a more comprehensive picture of these disorders.

Table 7.1. **Over-represented gene ontology (GO) biological process (BP) terms among proteins investigated in (A) Chapters 3-4 (NESDA cohort) and (B) Chapter 5 (schizophrenia cohorts).** GO BP enrichment analysis was carried out as described in **Chapter 2.** Only the top 20 terms are shown.

(A) GO BP ID	**Term**	**Expected genes**	**Observed genes**	**P-value**
GO:0051674	localization of cell	10.7	78	7.E-49
GO:2000145	regulation of cell motility	5.3	50	2.E-35
GO:0006950	response to stress	25.7	89	6.E-33
GO:0042330	taxis	6.3	48	1.E-29
GO:0006954	inflammatory response	3.2	37	2.E-29
GO:0051272	positive regulation of cellular component movement	2.8	34	2.E-27
GO:0010562	positive regulation of phosphorus metabolic process	8.4	50	1.E-25
GO:1903034	regulation of response to wounding	3.4	34	1.E-24
GO:0001932	regulation of protein phosphorylation	9.3	50	9.E-24
GO:0032501	multicellular organismal process	50.9	107	7.E-23
GO:0042060	wound healing	6.5	42	1.E-22
GO:0032101	regulation of response to external stimulus	2.9	30	2.E-22
GO:0031401	positive regulation of protein modification process	8.3	46	3.E-22
GO:0022610	biological adhesion	10.4	49	8.E-21
GO:0032879	regulation of localization	8.3	43	5.E-20
GO:0051247	positive regulation of protein metabolic process	9.0	45	5.E-20
GO:0001568	blood vessel development	5.2	35	2.E-19
GO:0009967	positive regulation of signal transduction	10.0	46	4.E-19
GO:0050790	regulation of catalytic activity	18.6	62	5.E-19
GO:1902531	regulation of intracellular signal transduction	14.1	54	7.E-19

(B) GO BP ID	**Term**	**Expected genes**	**Observed genes**	**P-value**
GO:0048584	positive regulation of response to stimulus	8.1	45	5.E-24
GO:0032101	regulation of response to external stimulus	3.1	31	5.E-23
GO:0060326	cell chemotaxis	1.1	21	1.E-21
GO:2000145	regulation of cell motility	2.9	27	2.E-19
GO:0042327	positive regulation of phosphorylation	4.1	30	1.E-18
GO:0043207	response to external biotic stimulus	3.7	28	1.E-17
GO:0006950	response to stress	14.4	48	2.E-17
GO:0042330	taxis	3.4	27	2.E-17
GO:0023056	positive regulation of signaling	5.4	32	2.E-17
GO:0010647	positive regulation of cell communication	5.4	32	3.E-17
GO:0006955	immune response	1.1	18	3.E-17
GO:0010562	positive regulation of phosphorus metabolic process	4.6	30	4.E-17
GO:0051272	positive regulation of cellular component movement	1.6	20	6.E-17
GO:1903034	regulation of response to wounding	1.9	21	9.E-17
GO:0018212	peptidyl-tyrosine modification	1.7	20	2.E-16
GO:0031401	positive regulation of protein modification process	4.6	29	3.E-16
GO:0009966	regulation of signal transduction	12.7	45	4.E-16
GO:0001932	regulation of protein phosphorylation	5.1	30	6.E-16
GO:0002544	chronic inflammatory response	0.1	9	7.E-16
GO:0009617	response to bacterium	1.1	17	1.E-15

Fundamental problems associated with research in mental disorders include lack of direct access to brain tissue, diagnostic criteria and patient heterogeneity, and the observational nature of the studies conducted. Ethical considerations prevent many experiments from being

conducted on human subjects and consequently human studies are largely observational. Most of these investigations, including those conducted in this thesis, assess patients already affected by the disorder. Furthermore, only peripheral tissues are easily accessible from patients before death, while brain tissue evaluated after death is often confounded by decades of psychiatric medication use, course of illness, suicide, and substance use, and sample sizes are often very limited. These factors limit the conclusions that can be drawn from the serum and PFC investigations. Future studies will be needed to confirm the results of this work and determine whether findings are due to pathophysiological or adaptive mechanisms in the disorders or a confounding/gender-associated factor. Evidence from animal and cellular models, neuroimaging studies, clinical trials, cognitive science, and further large-scale prospective and/or longitudinal peripheral and central omics studies will be needed to elucidate male and female pathophysiologies in these disorders. Finally, several issues with the diagnostic criteria used to classify patients and controls in mental disorders remain unresolved (see also **Chapter 1**). Different diagnosing physicians may be a source of heterogeneity across cohorts in **Chapters 5-6**. More fundamentally, however, the use of a dimensional approach in psychiatric research as opposed to the categorical one used in this work may be a more fruitful line of inquiry into the biological underpinnings of mental disorders.

7.3 Conclusions

The work carried out in this thesis represents the first large scale molecular profiling investigations of serum and brain tissue to comprehensively examine sex and female hormonal status in controls, schizophrenia, and MDD patients. Serum concentrations of a number of analytes involved in endocrine, metabolic, immune, and other processes varied with sex and female hormonal status, many of which were previously been found altered in schizophrenia and MDD. These processes may play a role in changing susceptibility and other characteristics of mental disorders known to differ between males and females. Male-specific elevations in immune and inflammatory molecules were associated with MDD and schizophrenia and these may have implications for immune hypotheses of these disorders. Furthermore, alterations of the HPG axis in schizophrenia and MDD suggested a potential role for elevated testosterone in females with these disorders. However, an investigation of sex-dependent differences in gene expression in the PFC of schizophrenia patients was limited by the small female sample size and proved inconclusive. Nevertheless, detection of differentially expressed genes was improved using robust regression and array weights and these were enriched in biological processes involved in cell signalling, neuronal apoptosis and mitochondrial functioning, response to estradiol, and others. Based on results of these investigations, future biomarker studies should take sex and female hormonal status into account during design and analysis

stages to improve reproducibility of these studies in mental disorders. As a whole these findings represent a substantial step forward in addressing the potential role of sex in the biological underpinnings of schizophrenia and MDD, which may help to explain differences between males and females in incidence, prevalence, age of onset, and other characteristics in these disorders. These results may also be useful in developing diagnostic technologies and directing sex-specific therapies to improve symptoms for patients with these disorders. However, a considerable amount of further work remains to be done to understand the aetiologies and pathophysiologies of schizophrenia, MDD, and other mental disorders in males and females.

Appendix

Table A.1. Multiplex immunoassay analytes examined in Chapters 3-5. For each assay, inclusion in the NESDA and combined schizophrenia cohorts is reported, together with percentage of missing values.

Analyte name	Short name	UniProtKB accession/ PubChem compound identifier	Chromosome	Chapter 5 (SCZ)	% missing (SCZ)	Chapters 3-4 (NESDA)	% missing (NESDA)
6Ckine		O00585	9			✓	0
α-1-Antichymotrypsin	AACT	P01011	14			✓	0
α-1-Antitrypsin	AAT	P01009	14	✓	0	✓	0.1
α-1-Microglobulin	A1Micro	P02760	9			✓	0
α-2-Macroglobulin	A2Macro	P01023	12	✓	0	✓	0.2
α-Fetoprotein		P02771	4	✓	7.1		
β-2-Microglobulin	B2M	P61769	15	✓	0	✓	0.2
Adiponectin		Q15848	3	✓	0	✓	0.2
Aldose Reductase		P15121	7			✓	0.3
Angiogenin		P03950	14			✓	0.2
Angiopoietin-2	ANG-2	O15123	8	✓	0.3	✓	0.1
Angiotensin-Converting Enzyme	ACE	P12821	17	✓	0	✓	0
Angiotensinogen		P01019	1	✓	3.9	✓	0.1
Apolipoprotein A-I	Apo A-I	P02647	11	✓	0	✓	0
Apolipoprotein A-II	Apo A-II	P02652	1			✓	0
Apolipoprotein A-IV	Apo A-IV	P06727	11			✓	0
Apolipoprotein B	Apo B	P04114	2			✓	0
Apolipoprotein C-I	Apo C-I	P02654	19			✓	0
Apolipoprotein C-III	Apo C-III	P02656	11	✓	0	✓	0
Apolipoprotein D	Apo D	P05090	3			✓	7.5
Apolipoprotein E	Apo E	P02649	19			✓	0.5
Apolipoprotein H	Apo H	P02749	17	✓	0	✓	0.2
Apolipoprotein(a)	Lp(a)	P08519	6	✓	0	✓	0.1
AXL Receptor Tyrosine Kinase	AXL	P30530	19	✓	0	✓	0
B cell-activating factor	BAFF	Q9Y275	13			✓	0
B Lymphocyte Chemoattractant	BLC	O43927	4	✓	15.6		
Brain Derived Neurotrophic Factor	BDNF	P23560	11	✓	0	✓	0.1
Cancer Antigen 15-3	CA 15-3	P15941	1			✓	0.1
Cancer Antigen 19-9	CA 19-9	Q9BXJ9	4	✓	7.8		
Carcinoembryonic Antigen	CEA	P06731	19	✓	2.3	✓	0.2
Cathepsin D		P07339	11			✓	0.2
CD40 antigen	CD40	Q6P2H9	20	✓	0.3	✓	0.1
CD40 Ligand	CD40L	P29965	X	✓	0.3	✓	0.4
CD5 Antigen-like	CD5L	O43866	1			✓	0.1
Cellular Fibronectin	cFib	P02751	2			✓	28.7
Chemokine CC-4	HCC-4	O15467	17	✓	0	✓	0
Chromogranin A	CgA	P10645	14	✓	2.3	✓	2.2
Clusterin	CLU	P10909	8			✓	0.1
Collagen IV		P02462 (a1), P08572 (a2), Q01955 (a3), P53420 (a4), P29400 (a5), Q14031 (a6)	13/2 /X			✓	0
Complement C3	C3	P01024	19	✓	0	✓	0.1
Complement Factor H – Related Protein 1	CFHR1	Q03591	1			✓	0
Cortisol		5754		✓	1.3	✓	0.2
C-Peptide		P01308	11			✓	0.3
C-Reactive Protein	CRP	P02741	1	✓	0.3	✓	1.3
Creatine Kinase-MB	CK-MB	P12277, P06732	14/ 19	✓	0	✓	6.1
Cystatin-C		P01034	20			✓	0
Endoglin		P17813	9			✓	0
Endostatin		P39060	21			✓	0.1
EN-RAGE		P80511	1	✓	0	✓	0
Eotaxin-1		P51671	17	✓	0.6	✓	11.6
Eotaxin-2		O00175	7			✓	0.3
Epidermal Growth Factor	EGF	P01133	4	✓	1.6	✓	0.3
Epidermal Growth Factor Receptor	EGFR	P00533	7	✓	0	✓	0.1
Epithelial-Derived Neutrophil-Activating Protein 78	ENA-78	P42830	4	✓	0	✓	0

E-Selectin		P16581	1			✓	0.1
Factor VII		P08709	13	✓	0	✓	0.1
FASLG Receptor	FAS	P25445	10	✓	0	✓	4.1
Fatty Acid-Binding Protein, adipocyte	FABP, adipocyte	P15090	8			✓	0.1
Ferritin	FRTN	P02794, P02792	11/ 19	✓	0	✓	2.2
Fetuin-A		P02765	3			✓	0
Fibulin-1C	Fib1C	P23142	22			✓	0.4
Follicle-Stimulating Hormone	FSH	P01225, P01215	11/6	✓	4.2	✓	2.0
Galectin-3		P17931	14			✓	0.1
Gelsolin		P06396	9			✓	0.2
Glucose-6-phosphate Isomerase	G6PI	P06744	19			✓	0
Glutathione S-Transferase α	GSTα/GSTA	P08263	6	✓	1.6	✓	26.1
Granulocyte Colony-Stimulating Factor	G-CSF	P09919	17	✓	5.8		
Growth Hormone	GH	P01241	17	✓	6.8	✓	15.5
Growth-Regulated α protein	GRO-α/GRO-A	P09341	4	✓	0	✓	0
Haptoglobin		P00738	16	✓	0.6	✓	4.5
Heparin-Binding EGF-Like Growth Factor	HB-EGF	Q99075	5	✓	0		
Hepatocyte Growth Factor	HGF	P14210	7	✓	0.3	✓	0.3
Hepatocyte Growth Factor receptor	HGF receptor	P08581	7			✓	0
Hepsin		P05981	19			✓	0
Human Epidermal Growth Factor Receptor 2	HER-2	P04626	17			✓	0
Immunoglobulin A	IgA			✓	0	✓	0.1
Immunoglobulin E	IgE			✓	14.3		
Immunoglobulin M	IgM			✓	0	✓	0.2
Insulin		P01308	11	✓	1.0	✓	20.6
Insulin-like Growth Factor Binding Protein 1	IGFBP-1	P08833	7			✓	1.4
Insulin-like Growth Factor Binding Protein 2	IGFBP-2	P18065	2	✓	0	✓	0
Insulin-like Growth Factor Binding Protein 3	IGFBP-3	P17936	7			✓	0.1
Insulin-like Growth Factor Binding Protein 4	IGFBP-4	P22692	17			✓	0.1
Insulin-like Growth Factor Binding Protein 5	IGFBP-5	P24593	2			✓	0.1
Insulin-like Growth Factor Binding Protein 6	IGFBP-6	P24592	12			✓	0.1
Intercellular Adhesion Molecule 1	ICAM-1	P05362	19	✓	0	✓	0.1
Interferon γ Induced Protein 10	IP-10	P02778	4			✓	0.6
Interleukin-1 receptor antagonist	IL-1ra	P18510	2	✓	1.0	✓	1.6
Interleukin-2 receptor α	IL-2 receptor α/ IL-2RA	P01589	10			✓	0
Interleukin-3	IL-3	P08700	5	✓	6.8		
Interleukin-6 receptor	IL-6r	P08887	1			✓	0.1
Interleukin-6 receptor subunit β	IL-6rβ/IL-6rB	P40189	5			✓	0.2
Interleukin-7	IL-7	P13232	8	✓	1.9		
Interleukin-8	IL-8	P10145	4	✓	2.9	✓	13.1
Interleukin-10	IL-10	P22301	1	✓	0		
Interleukin-12 Subunit p40	IL-12p40	P29460	5			✓	3.5
Interleukin-13	IL-13	P35225	5	✓	2.9		
Interleukin-15	IL-15	P40933	4	✓	5.2		
Interleukin-16	IL-16	Q14005	15	✓	0	✓	0.2
Interleukin-18	IL-18	Q14116	11	✓	0	✓	0.7
Interleukin-23	IL-23	Q9NPF7	12			✓	4.8
Kallikrein 5		Q9Y337	19			✓	0.8
Lactoylglutathione lyase	LGL	Q04760	6			✓	0.1
Latency-Associated Peptide of Transforming Growth Factor β 1	LAP TGF-b1	P01137	19			✓	0
Lectin-Like Oxidized LDL Receptor 1	LOX-1	P78380	12			✓	28.7
Leptin		P41159	7	✓	0	✓	0.2
Luteinizing Hormone	LH	P01229, P01215	19/6	✓	1.9	✓	28.4
Macrophage Colony-Stimulating Factor 1	M-CSF	P09603	1			✓	3.8
Macrophage Derived Chemokine	MDC	O00626	16	✓	0	✓	0
Macrophage Inflammatory Protein-1 α	MIP-1α/MIP-1A	P10147	17	✓	0.3		
Macrophage Inflammatory Protein-1 β	MIP-1β/MIP-1B	P13236	17	✓	0	✓	0.1
Macrophage Inflammatory Protein-3 β	MIP-3β/MIP-3B	Q99731	9			✓	0
Macrophage Migration Inhibitory Factor	MIF	P14174	22	✓	0.6	✓	0.1
Macrophage Stimulating Protein	MSP	P26927	3			✓	0.2
Matrix Metalloproteinase-1	MMP-1	P03956	11			✓	0.9
Matrix Metalloproteinase-3	MMP-3	P08254	11	✓	0	✓	0.1
Matrix Metalloproteinase-7	MMP-7	P09237	11			✓	0.2
Matrix Metalloproteinase-9 (total)	MMP-9	P14780	20			✓	0.2
Matrix Metalloproteinase-10	MMP-10	P09238	11			✓	0.2
Mesothelin	MSLN	Q13421	16			✓	0.1
Monocyte Chemotactic Protein 1	MCP-1	P13500	17	✓	0	✓	0.3
Monocyte Chemotactic Protein 2	MCP-2	P80075	17			✓	0.4
Monocyte Chemotactic Protein 4	MCP-4	Q99616	17			✓	2.0

Monokine Induced by γ Interferon	MIG	Q07325	4			✓	0.3
Myeloid Progenitor Inhibitory Factor 1	MPIF-1	P55773	17			✓	0.2
Myeloperoxidase	MPO	P05164	17	✓	0	✓	1.7
Myoglobin		P02144	22	✓	0	✓	0.2
Neuron Specific Enolase	NSE	P09104	12			✓	0
Neuronal Cell Adhesion Molecule	NrCAM	Q92823	7			✓	2.0
Neuropilin-1		O14786	10			✓	0
Neutrophil Gelatinase Associated Lipocalin	NGAL	P80188	9			✓	0
N-terminal prohormone of brain natriuretic peptide	NT proBNP	P16860	1			✓	6.3
Osteopontin		P10451	4			✓	16.3
Osteoprotegerin	OPG	O00300	8			✓	0
Pancreatic Polypeptide	PPP	P01298	17	✓	0	✓	0.1
Pepsinogen I	PGI	O95576	11			✓	0.2
Phosphoserine Aminotransferase	PSAT	Q9Y617	9			✓	0.1
Plasminogen Activator Inhibitor 1	PAI-1	P05121	7	✓	0	✓	0.2
Platelet-Derived Growth Factor BB	PDGF-BB	P01127	22	✓	0.3	✓	0.1
Progesterone		5994		✓	1.3	✓	19.9
Prolactin	PRL	P01236	6	✓	0	✓	0
Prostasin		Q16651	16			✓	0.1
Prostatic Acid Phosphatase		P15309	3	✓	0		
Pulmonary and Activation-Regulated Chemokine	PARC	P55774	17	✓	0.3	✓	0
Receptor for advanced glycosylation end products	RAGE	Q15109	6	✓	0	✓	0.1
Receptor tyrosine-protein kinase erbB-3	ErbB3	P21860	12			✓	0.1
Resistin		Q9HD89	19	✓	0	✓	0.1
Serotransferrin	Transferrin	P02787	3			✓	0
Serum Amyloid P-Component	SAP	P02743	1	✓	0	✓	0.1
Serum Glutamic Oxaloacetic Transaminase	SGOT	P17174	10	✓	0		
Sex Hormone-Binding Globulin	SHBG	P04278	17	✓	0	✓	0.4
Sortilin		Q99523	1	✓	0	✓	0.1
Stem Cell Factor	SCF	P21583	12	✓	0	✓	0.5
Stromal cell derived factor 1	SDF-1	P48061	10			✓	0
Superoxide Dismutase 1, soluble	SOD-1	P00441	21	✓	0	✓	0.8
Tamm-Horsfall Urinary Glycoprotein	THP	P07911	16			✓	0
T-Cell-Specific Protein RANTES	RANTES	P13501	17	✓	0	✓	0.3
Tenascin-C	TN-C	P24821	9	✓	0	✓	0.1
Testosterone (total)		6013		✓	0.3	✓	10.0
Tetranectin		P05452	3			✓	0.2
Thrombomodulin	TM	P07204	20			✓	0
Thrombopoietin		P40225	3	✓	0		
Thrombospondin-1		P07996	15	✓	0	✓	0
Thyroglobulin	TG	P01266	8			✓	25.5
Thyroid Stimulating Hormone	TSH	P01215, P01222	6/1	✓	0	✓	0.4
Thyroxine-Binding Globulin	TBG	P05543	X	✓	0	✓	0.2
Tissue Inhibitor of Metalloproteinases 1	TIMP-1	P01033	X	✓	0	✓	0.2
Tissue type Plasminogen activator	tPA	P00750	8			✓	2.8
TNF-Related Apoptosis-Inducing Ligand Receptor 3	TRAIL-R3	O14798	8	✓	0	✓	0
Transthyretin	TTR	P02766	18			✓	0
Trefoil Factor 3	TFF3	Q07654	21			✓	0
Tumor necrosis factor receptor 2	TNFR2	P20333	1	✓	0	✓	0.2
Tumor Necrosis Factor Receptor I	TNFRI	P19438	12			✓	0
Tumor Necrosis Factor α	TNF α	P01375	6	✓	10.4		
Tyrosine kinase with Ig and EGF homology domains 2	TIE-2	Q02763	9			✓	0.1
Urokinase-type Plasminogen Activator	uPA	P00749	10			✓	0.1
Urokinase-type plasminogen activator receptor	uPAR	Q03405	19			✓	0.6
Vascular Cell Adhesion Molecule 1	VCAM-1	P19320	1	✓	0	✓	0.2
Vascular Endothelial Growth Factor	VEGF	P15692	6	✓	0	✓	0.1
Vascular Endothelial Growth Factor C	VEGF-C	P49767	4			✓	0.1
Vascular Endothelial Growth Factor Receptor 2	VEGFR-2	P35968	4			✓	0
Vascular endothelial growth factor Receptor 3	VEGFR-3	P35916	5			✓	1.0
Vitamin D-Binding Protein	VDBP	P02774	4			✓	0.1
Vitamin K-Dependent Protein S	VKDPS	P07225	3			✓	0
Vitronectin		P04004	17			✓	0.1
von Willebrand Factor	vWF	P04275	12	✓	0	✓	1.1
YKL-40		P36222	1			✓	0.2

References

1. American Psychiatric Association, *Diagnostic and statistical manual of mental disorders : DSM-5.* (Author, Washington, DC, ed. 5th, 2013).

2. A. Metastasio, S. Bahn, Utility of biomarkers to improve the diagnosis and treatment of schizophrenia. *Future Neurol.* **3**, 619–622 (2008).

3. World Economic Forum, Harvard School of Public Health, “The Global Economic Burden of Non-communicable Diseases” (Geneva, 2011).

4. C. E. Bird, P. P. Rieker, Gender matters: an integrated model for understanding men’s and women's health. *Soc. Sci. Med.* **48**, 745–755 (1999).

5. C. Vlassoff, C. Garcia Moreno, Placing gender at the centre of health programming: challenges and limitations. *Soc. Sci. Med.* **54**, 1713–23 (2002).

6. J. L. Johnson, L. Greaves, R. Repta, Better science with sex and gender: Facilitating the use of a sex and gender-based analysis in health research. *Int. J. Equity Health.* **8**, 14 (2009).

7. V. Regitz-Zagrosek, Sex and gender differences in health. Science & Society Series on Sex and Science. *EMBO Rep.* **13**, 596–603 (2012).

8. Institute of Medicine (U.S.) Committee on Understanding the Biology of Sex and Gender Differences, *Exploring the Biological Contributions to Human Health: Does Sex Matter?* (National Academies Press, Washington, DC, 2001).

9. M. J. H. Begemann, C. F. Dekker, M. van Lunenburg, I. E. Sommer, Estrogen augmentation in schizophrenia: a quantitative review of current evidence. *Schizophr. Res.* **141**, 179–84 (2012).

10. E. Schwarz *et al.*, Sex-specific serum biomarker patterns in adults with Asperger’s syndrome. *Mol. Psychiatry.* **16**, 1213–20 (2011).

11. E. Fombonne, The Changing Epidemiology of Autism. *J. Appl. Res. Intellect. Disabil.* **18**, 281–294 (2005).

12. K. M. Abel, R. Drake, J. M. Goldstein, Sex differences in schizophrenia. *Int. Rev. Psychiatry.* **22**, 417–28 (2010).

13. E. Bromet *et al.*, Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med.* **9**, 90 (2011).

14. K. Hawton, Sex and suicide: Gender differences in suicidal behaviour. *Br. J. Psychiatry.* **177**, 484–485 (2000).

15. D. F. Tolin, E. B. Foa, Sex differences in trauma and posttraumatic stress disorder: a quantitative review of 25 years of research. *Psychol. Bull.* **132**, 959–92 (2006).

16. M. A. de Mathis *et al.*, Gender differences in obsessive-compulsive disorder: a literature review. *Rev. Bras. Psiquiatr.* **33**, 390–399 (2011).

17. I. Kawa *et al.*, Gender differences in bipolar disorder: age of onset, course, comorbidity, and symptom presentation. *Bipolar Disord.* **7**, 119–25 (2005).

18. J. I. Hudson, E. Hiripi, H. G. Pope, R. C. Kessler, The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biol. Psychiatry*. **61**, 348–58 (2007).

19. J. B. Becker, M. Hu, Sex differences in drug abuse. *Front. Neuroendocrinol.* **29**, 36–47 (2008).

20. E. M. Cale, S. O. Lilienfeld, Sex differences in psychopathy and antisocial personality disorder. *Clin. Psychol. Rev.* **22**, 1179–1207 (2002).

21. R. A. Sansone, L. A. Sansone, Gender patterns in borderline personality disorder. *Innov. Clin. Neurosci.* **8**, 16–20 (2011).

22. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV* (Author, Washington, DC, ed. 4th, 2000).

23. E. M. Corbitt, T. A. Widiger, Sex Differences Among the Personality Disorders: An Exploration of the Data. *Clin. Psychol. Sci. Pract.* **2**, 225–238 (1995).

24. D. E. Pankevich, T. Wizemann, B. M. Altevogt, in *Institute of Medicine (US) Forum on Neuroscience and Nervous System Disorders* (The National Academies Press, Washington, DC, 2011).

25. A. P. Arnold, The organizational-activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Horm. Behav.* **55**, 570–8 (2009).

26. S. Hiller-Sturmhöfel, A. Bartke, The endocrine system: an overview. *Alcohol Health Res. World*. **22**, 153–64 (1998).

27. T. Simoncini, A. R. Genazzani, Non-genomic actions of sex steroid hormones. *Eur. J. Endocrinol.* **148**, 281–92 (2003).

28. R. G. Brzyski, J. Knudtson, in *The Merck Manual Professional Edition Online*, R. S. Porter, J. L. Kaplan, Eds. (Merck & Co, Whitehouse Station, NJ, 2015).

29. V. Viau, Functional cross-talk between the hypothalamic-pituitary-gonadal and -adrenal axes. *J. Neuroendocrinol.* **14**, 506–13 (2002).

30. Y. M. Ulrich-Lai, J. P. Herman, Neural regulation of endocrine and autonomic stress responses. *Nat. Rev. Neurosci.* **10**, 397–409 (2009).

31. C. M. Pariante, S. L. Lightman, The HPA axis in major depression: classical theories and new developments. *Trends Neurosci.* **31**, 464–8 (2008).

32. E. G. Spratt *et al.*, Enhanced cortisol response to stress in children in autism. *J. Autism Dev. Disord.* **42**, 75–81 (2012).

33. C. Daban, E. Vieta, P. Mackin, A. H. Young, Hypothalamic-pituitary-adrenal axis and bipolar disorder. *Psychiatr. Clin. North Am.* **28**, 469–80 (2005).

34. A. J. Bradley, T. G. Dinan, A systematic review of hypothalamic-pituitary-adrenal axis function in schizophrenia: implications for mortality. *J. Psychopharmacol.* **24**, 91–118 (2010).

35. F. Tanriverdi, L. F. G. Silveira, G. S. MacColl, P. M. G. Bouloux, The hypothalamic-pituitary-gonadal axis: immune function and autoimmunity. *J. Endocrinol.* **176**, 293–304 (2003).

36. R. Pasquali, The hypothalamic-pituitary-adrenal axis and sex hormones in chronic stress and obesity: pathophysiological and clinical aspects. *Ann. N. Y. Acad. Sci.* **1264**, 20–35 (2012).

37. L. I. McKay, J. A. Cidlowski, in *Holland-Frei Cancer Medicine*, D. Kufe, R. Pollock, R. Weichselbaum, Eds. (BC Decker, Hamilton (ON), ed. 6, 2003; http://www.ncbi.nlm.nih.gov/books/NBK13780/).

38. L. Cahill, Why sex matters for neuroscience. *Nat Rev Neurosci.* **7**, 477–84 (2006).

39. W. C. J. Chung, A. P. Auger, Gender differences in neurodevelopment and epigenetics. *Pflugers Arch.* **465**, 573–84 (2013).

40. A. Gogos, P. Kwek, M. van den Buuse, The role of estrogen and testosterone in female rats in behavioral models of relevance to schizophrenia. *Psychopharmacology (Berl).* **219**, 213–24 (2012).

41. M. Cyr, F. Calon, M. Morissette, T. Di Paolo, Estrogenic modulation of brain activity: implications for schizophrenia and Parkinson's disease. *J. Psychiatry Neurosci.* **27**, 12–27 (2002).

42. S. Nishizawa *et al.*, Differences between males and females in rates of serotonin synthesis in human brain. *Proc. Natl. Acad. Sci.* **94**, 5308–5313 (1997).

43. M. Arad, I. Weiner, Contrasting effects of increased and decreased dopamine transmission on latent inhibition in ovariectomized rats and their modulation by 17beta-estradiol: an animal model of menopausal psychosis? *Neuropsychopharmacology*. **35**, 1570–82 (2010).

44. E. Jazin, L. Cahill, Sex differences in molecular neuroscience: from fruit flies to humans. *Nat Rev Neurosci.* **11**, 9–17 (2010).

45. D. H. Skuse, X-linked genes and mental functioning. *Hum. Mol. Genet.* **14 Spec No**, R27–32 (2005).

46. H. A. Lubs, R. E. Stevenson, C. E. Schwartz, Fragile X and X-linked intellectual disability: four decades of discovery. *Am. J. Hum. Genet.* **90**, 579–90 (2012).

47. S. Najjar, D. M. Pearlman, K. Alper, A. Najjar, O. Devinsky, Neuroinflammation and psychiatric illness. *J. Neuroinflammation*. **10**, 43 (2013).

48. A. Rozanski, J. A. Blumenthal, J. Kaplan, Impact of Psychological Factors on the Pathogenesis of Cardiovascular Disease and Implications for Therapy. *Circulation*. **99**, 2192–2217 (1999).

49. J. W. Newcomer, Metabolic syndrome and mental illness. *Am. J. Manag. Care*. **13**, S170–7 (2007).

50. B. Rudisch, C. B. Nemeroff, Epidemiology of comorbid coronary artery disease and depression. *Biol. Psychiatry*. **54**, 227–240 (2003).

51. S. Prabakaran *et al.*, Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress. *Mol. Psychiatry*. **9**, 684–97, 643 (2004).

52. R. J. Anderson, K. E. Freedland, R. E. Clouse, P. J. Lustman, The Prevalence of Comorbid Depression in Adults With Diabetes: A meta-analysis. *Diabetes Care*. **24**, 1069–1078 (2001).

53. E. N. Fish, The X-files in immunity: Sex-based differences predispose immune responses. *Nat Rev Immunol*. **8**, 737–744 (2008).

54. K. M. Lenz, M. M. McCarthy, A starring role for microglia in brain sex differences. *Neuroscientist*. **21**, 306–21 (2014).

55. A. J. Bruce-Keller *et al.*, Antiinflammatory effects of estrogen on microglial activation. *Endocrinology*. **141**, 3646–56 (2013).

56. E. Vegeto *et al.*, Estrogen receptor-alpha mediates the brain antiinflammatory activity of estradiol. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9614–9 (2003).

57. M. L. Power, J. Schulkin, Sex differences in fat storage, fat metabolism, and the health risks from obesity: possible evolutionary origins. *Br. J. Nutr.* **99**, 931–40 (2008).

58. M. Holzenberger *et al.*, IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature*. **421**, 182–7 (2003).

59. C. Borrás *et al.*, Mitochondria from females exhibit higher antioxidant gene expression and lower oxidative damage than males. *Free Radic. Biol. Med.* **34**, 546–552 (2003).

60. E. Blaak, Gender differences in fat metabolism. *Curr. Opin. Clin. Nutr. Metab. Care*. **4**, 499–502 (2001).

61. I. F. Godsland *et al.*, The effects of different formulations of oral contraceptive agents on lipid and carbohydrate metabolism. *N. Engl. J. Med.* **323**, 1375–81 (1990).

62. R. Gur *et al.*, Sex differences in regional cerebral glucose metabolism during a resting state. *Science (80-.).* **267**, 528–531 (1995).

63. D. G. Stein, Brain damage, sex hormones and recovery: a new role for progesterone and estrogen? *Trends Neurosci.* **24**, 386–391 (2001).

64. M. Gandhi, F. Aweeka, R. M. Greenblatt, T. F. Blaschke, Sex differences in pharmacokinetics and pharmacodynamics. *Annu. Rev. Pharmacol. Toxicol.* **44**, 499–523 (2004).

65. R. A. M. Haast, D. R. Gustafson, A. J. Kiliaan, Sex differences in stroke. *J. Cereb. Blood Flow Metab.* **32**, 2100–7 (2012).

66. W. G. for the W. H. I. Investigators, Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results From the Women's Health Initiative Randomized Controlled Trial. *JAMA*. **288**, 321–333 (2002).

67. J.-P. Baillargeon, D. K. McClish, P. A. Essah, J. E. Nestler, Association between the current use of low-dose oral contraceptives and cardiovascular arterial disease: a meta-analysis. *J. Clin. Endocrinol. Metab.* **90**, 3863–70 (2005).

68. J. Vandenbroucke, Jan P. Rosing, S. Bloemenkamp, Kitty W.M. Middeldorp, F. M. Helmerhorst, B. N. Bouma, F. R. Rosendaal, Oral Contraceptives and the Risk of Venous Thrombosis. *New Engl J Med*. **344**, 1527–1535 (2001).

69. M. Mendelsohn, R. Karas, Molecular and cellular basis of cardiovascular gender differences. *Science (80-.).* **308**, 1583–1587 (2005).

70. D. W. Schwertz, S. Penckofer, Sex differences and the effects of sex hormones on hemostasis and vascular reactivity. *Heart Lung*. **30**, 401–28 (2001).

71. D. W. Losordo, J. M. Isner, Estrogen and Angiogenesis : A Review. *Arterioscler. Thromb. Vasc. Biol.* **21**, 6–12 (2001).

72. L. D. McCullough, P. D. Hurn, Estrogen and ischemic neuroprotection: an integrated view. *Trends Endocrinol. Metab.* **14**, 228–235 (2003).

73. E. Murphy, Estrogen signaling and cardiovascular disease. *Circ. Res.* **109**, 687–96 (2011).

74. World Economic Forum, "The Global Gender Gap Report 2013" (Geneva, 2013), (available at http://www3.weforum.org/docs/WEF_GenderGap_Report_2013.pdf).

75. C. P. Cross, D.-L. M. Cyrenne, G. R. Brown, Sex differences in sensation-seeking: a meta-analysis. *Sci. Rep.* **3**, 2486 (2013).

76. P. M. Galdas, F. Cheater, P. Marshall, Men and health help-seeking behaviour: literature review. *J. Adv. Nurs.* **49**, 616–23 (2005).

77. N. Breslau, Gender differences in trauma and posttraumatic stress disorder. *J. Gend. Specif. Med.* **5**, 34–40 (2002).

78. M. P. Matud, Gender differences in stress and coping styles. *Pers. Individ. Dif.* **37**, 1401–1415 (2004).

79. P. T. Costa, A. Terracciano, R. R. McCrae, Gender differences in personality traits across cultures: robust and surprising findings. *J. Pers. Soc. Psychol.* **81**, 322–31 (2001).

80. M. Denton, V. Walters, Gender differences in structural and behavioral determinants of health: an analysis of the social production of health. *Soc. Sci. Med.* **48**, 1221–1235 (1999).

81. G. M. Reed, J. M. Correia, P. Esparza, S. Saxena, M. Maj, The WPA-WHO Global Survey of Psychiatrists' Attitudes Towards Mental Disorders Classification. *World Psychiatry*. **10**, 118–131 (2011).

82. H.-U. Wittchen, Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review. *J. Psychiatr. Res.* **28**, 57–84 (1994).

83. T. A. Widiger, D. B. Samuel, Diagnostic categories or dimensions? A question for the Diagnostic and statistical manual of mental disorders--fifth edition. *J. Abnorm. Psychol.* **114**, 494–504 (2005).

84. T. Insel *et al.*, Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry*. **167**, 748–51 (2010).

85. C. A. Sanislow *et al.*, Developing constructs for psychopathology research: research domain criteria. *J. Abnorm. Psychol.* **119**, 631–9 (2010).

86. P. D. White, H. Rickards, A. Z. J. Zeman, Time to end the distinction between mental and neurological illnesses. *BMJ*. **344**, e3454 (2012).

87. J. C. Masdeu, Neuroimaging in psychiatric disorders. *Neurotherapeutics*. **8**, 93–102 (2011).

88. K. Pennington, D. Cotter, M. J. Dunn, The role of proteomics in investigating psychiatric disorders. *Br. J. Psychiatry*. **187**, 4–6 (2005).

89. A. Hayashi-Takagi, M. P. Vawter, K. Iwamoto, Peripheral biomarkers revisited: integrative profiling of peripheral samples for psychiatric research. *Biol. Psychiatry*. **75**, 920–8 (2014).

90. K. J. Brennand *et al.*, Modelling schizophrenia using human induced pluripotent stem cells. *Nature*. **473**, 221–5 (2011).

91. M. K. Chan *et al.*, Applications of blood-based protein biomarker strategies in the study of psychiatric disorders. *Prog. Neurobiol.* **122C**, 45–72 (2014).

92. T. R. Insel, Rethinking schizophrenia. *Nature*. **468**, 187–93 (2010).

93. J. C. Ballenger *et al.*, Consensus statement on transcultural issues in depression and anxiety from the International Consensus Group on Depression and Anxiety. *J. Clin. Psychiatry*. **62 Suppl 1**, 47–55 (2001).

94. H.-U. Wittchen *et al.*, Generalized anxiety and depression in primary care: prevalence, recognition, and management. *J. Clin. Psychiatry*. **63 Suppl 8**, 24–34 (2002).

95. World Health Organization (WHO), "The global burden of disease: 2004 update" (2008), (available at http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/).

96. J. Olesen, A. Gustavsson, M. Svensson, H.-U. Wittchen, B. Jönsson, The economic cost of brain disorders in Europe. *Eur. J. Neurol.* **19**, 155–62 (2012).

97. B. F. Grant, T. C. Harford, Comorbidity between DSM-IV alcohol use disorders and major depression: results of a national survey. *Drug Alcohol Depend.* **39**, 197–206 (1995).

98. P. F. Buckley, B. J. Miller, D. S. Lehrer, D. J. Castle, Psychiatric comorbidities and schizophrenia. *Schizophr. Bull.* **35**, 383–402 (2009).

99. R. M. A. Hirschfeld, The Comorbidity of Major Depression and Anxiety Disorders: Recognition and Management in Primary Care. *Prim. Care Companion J. Clin. Psychiatry*. **3**, 244–254 (2001).

100. Y. Aina, J. L. Susman, Understanding comorbidity with depression and anxiety disorders. *J. Am. Osteopath. Assoc.* **106**, S9–14 (2006).

101. S. L. Burcusa, W. G. Iacono, Risk for recurrence in depression. *Clin. Psychol. Rev.* **27**, 959–85 (2007).

102. J. M. Bostwick, V. S. Pankratz, Affective disorders and suicide risk: a reexamination. *Am. J. Psychiatry*. **157**, 1925–32 (2000).

103. K. M. Holma *et al.*, Incidence and predictors of suicide attempts in DSM-IV major depressive disorder: a five-year prospective study. *Am. J. Psychiatry*. **167**, 801–8 (2010).

104. I. Kirsch *et al.*, Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med.* **5**, e45 (2008).

105. J. C. Fournier *et al.*, Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*. **303**, 47–53 (2010).

106. S. Pampallona, P. Bollini, G. Tibaldi, B. Kupelnick, C. Munizza, Combined pharmacotherapy and psychological treatment for depression: a systematic review. *Arch. Gen. Psychiatry*. **61**, 714–9 (2004).

107. M. S. George *et al.*, Vagus nerve stimulation: a new tool for brain research and therapy. *Biol. Psychiatry*. **47**, 287–295 (2000).

108. J. P. O'Reardon *et al.*, Efficacy and safety of transcranial magnetic stimulation in the acute treatment of major depression: a multisite randomized controlled trial. *Biol. Psychiatry*. **62**, 1208–16 (2007).

109. M. Olfson, S. Marcus, H. A. Sackeim, J. Thompson, H. A. Pincus, Use of ECT for the Inpatient Treatment of Recurrent Major Depression. *Am. J. Psychiatry*. **155**, 22–29 (1998).

110. D. L. Hudson, H. W. Neighbors, A. T. Geronimus, J. S. Jackson, The relationship between socioeconomic position and depression among a US nationally representative sample of African Americans. *Soc. Psychiatry Psychiatr. Epidemiol.* **47**, 373–81 (2012).

111. A. Simning, E. van Wijngaarden, Y. Conwell, Anxiety, mood, and substance use disorders in United States African-American public housing residents. *Soc. Psychiatry Psychiatr. Epidemiol.* **46**, 983–92 (2011).

112. B. P. Dohrenwend *et al.*, Socioeconomic status and psychiatric disorders: the causation-selection issue. *Science*. **255**, 946–52 (1992).

113. C. Muntaner, W. W. Eaton, C. Diala, R. C. Kessler, P. D. Sorlie, Social class, assets, organizational control and the prevalence of common groups of psychiatric disorders. *Soc. Sci. Med.* **47**, 2043–53 (1998).

114. P. F. Sullivan, M. C. Neale, K. S. Kendler, Genetic Epidemiology of Major Depression: Review and Meta-Analysis. *Am. J. Psychiatry*. **157**, 1552–1562 (2000).

115. E. J. Nestler *et al.*, Neurobiology of Depression. *Neuron*. **34**, 13–25 (2002).

116. O. Abe *et al.*, Voxel-based analyses of gray/white matter volume and diffusion tensor data in major depression. *Psychiatry Res.* **181**, 64–70 (2010).

117. G. Hasler, Pathophysiology of depression: do we have any solid evidence of interest to clinicians? *World Psychiatry*. **9**, 155–61 (2010).

118. Y. I. Sheline, P. W. Wang, M. H. Gado, J. G. Csernansky, M. W. Vannier, Hippocampal atrophy in recurrent major depression. *Proc. Natl. Acad. Sci.* **93**, 3908–3913 (1996).

119. G. Rajkowska *et al.*, Morphometric evidence for neuronal and glial prefrontal cell pathology in major depression. *Biol. Psychiatry*. **45**, 1085–1098 (1999).

120. D. Mehta, A. Menke, E. B. Binder, Gene expression studies in major depression. *Curr. Psychiatry Rep.* **12**, 135–44 (2010).

121. D. Martins-de-Souza *et al.*, Identification of proteomic signatures associated with depression and psychotic depression in post-mortem brains from major depression patients. *Transl. Psychiatry*. **2**, e87 (2012).

122. A. L. Lopresti, G. L. Maker, S. D. Hood, P. D. Drummond, A review of peripheral biomarkers in major depression: the potential of inflammatory and oxidative stress biomarkers. *Prog. Neuropsychopharmacol. Biol. Psychiatry*. **48**, 102–11 (2014).

123. Y. Dowlati *et al.*, A meta-analysis of cytokines in major depression. *Biol. Psychiatry*. **67**, 446–57 (2010).

124. M. B. Howren, D. M. Lamkin, J. Suls, Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. *Psychosom. Med.* **71**, 171–86 (2009).

125. S. A. Hiles, A. L. Baker, T. de Malmanche, J. Attia, Interleukin-6, C-reactive protein and interleukin-10 after antidepressant treatment in people with depression: a meta-analysis. *Psychol. Med.* **42**, 2015–26 (2012).

126. J. Hannestad, N. DellaGioia, M. Bloch, The effect of antidepressant medication treatment on serum levels of inflammatory cytokines: a meta-analysis. *Neuropsychopharmacology*. **36**, 2452–9 (2011).

127. J.-I. Iga *et al.*, Gene expression and association analysis of vascular endothelial growth factor in major depressive disorder. *Prog. Neuropsychopharmacol. Biol. Psychiatry*. **31**, 658–63 (2007).

128. S. Sen, R. Duman, G. Sanacora, Serum brain-derived neurotrophic factor, depression, and antidepressant medications: meta-analyses and implications. *Biol. Psychiatry*. **64**, 527–32 (2008).

129. V. Stelzhammer *et al.*, Proteomic changes in serum of first onset, antidepressant drug-naïve major depression patients. *Int. J. Neuropsychopharmacol.* **17**, 1599–608 (2014).

130. E. Domenici *et al.*, Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections. *PLoS One*. **5**, e9166 (2010).

131. G. I. Papakostas *et al.*, Assessment of a multi-assay, serum-based biological diagnostic test for major depressive disorder: a pilot and replication study. *Mol. Psychiatry*. **18**, 332–9 (2013).

132. R. Ramasubbu, Insulin resistance: a metabolic link between depressive disorder and atherosclerotic vascular diseases. *Med. Hypotheses*. **59**, 537–551 (2002).

133. K. S. Kendler, M. Gatz, C. O. Gardner, N. L. Pedersen, Personality and major depression: a Swedish longitudinal, population-based twin study. *Arch. Gen. Psychiatry*. **63**, 1113–20 (2006).

134. S. Ripke *et al.*, A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry*. **18**, 497–511 (2013).

135. J. Mill, A. Petronis, Molecular studies of major depressive disorder: the epigenetic perspective. *Mol. Psychiatry*. **12**, 799–814 (2007).

136. V. Krishnan, E. J. Nestler, The molecular neurobiology of depression. *Nature*. **455**, 894–902 (2008).

137. S. A. Vreeburg *et al.*, Major depressive disorder and hypothalamic-pituitary-adrenal axis activity: results from a large cohort study. *Arch. Gen. Psychiatry*. **66**, 617–26 (2009).

138. K. J. Parker, A. F. Schatzberg, D. M. Lyons, Neuroendocrine aspects of hypercortisolism in major depression. *Horm. Behav.* **43**, 60–6 (2003).

139. H. M. Burke, M. C. Davis, C. Otte, D. C. Mohr, Depression and cortisol responses to psychological stress: a meta-analysis. *Psychoneuroendocrinology*. **30**, 846–56 (2005).

140. E. S. Brown, F. P. Varghese, B. S. McEwen, Association of depression with medical illness: does cortisol play a role? *Biol. Psychiatry*. **55**, 1–9 (2004).

141. P. W. Gold, G. P. Chrousos, Organization of the stress system and its dysregulation in melancholic and atypical depression: high vs low CRH/NE states. *Mol. Psychiatry*. **7**, 254–75 (2002).

142. O. J. G. Schiepers, M. C. Wichers, M. Maes, Cytokines and major depression. *Prog. Neuropsychopharmacol. Biol. Psychiatry*. **29**, 201–17 (2005).

143. R. Dantzer, J. C. O'Connor, G. G. Freund, R. W. Johnson, K. W. Kelley, From inflammation to sickness and depression: when the immune system subjugates the brain. *Nat. Rev. Neurosci.* **9**, 46–56 (2008).

144. G. E. Miller, S. Cohen, A. K. Ritchey, Chronic psychological stress and the regulation of pro-inflammatory cytokines: A glucocorticoid-resistance model. *Heal. Psychol.* **21**, 531–41 (2002).

145. T. Kreisel *et al.*, Dynamic microglial alterations underlie stress-induced depressive-like behavior and suppressed neurogenesis. *Mol. Psychiatry*. **19**, 699–709 (2014).

146. G. S. Alexopoulos *et al.*, “Vascular depression” hypothesis. *Arch. Gen. Psychiatry*. **54**, 915–22 (1997).

147. Y. I. Sheline *et al.*, Support for the vascular depression hypothesis in late-life depression: results of a 2-site, prospective, antidepressant treatment trial. *Arch. Gen. Psychiatry*. **67**, 277–85 (2010).

148. R. Kessler, Epidemiology of women and depression. *J. Affect. Disord.* **74**, 5–13 (2003).

149. M. M. Weissman *et al.*, Cross-National Epidemiology of Major Depression and Bipolar Disorder. *JAMA*. **276**, 293–9 (1996).

150. B. L. Hankin, L. Y. Abramson, Development of gender differences in depression: an elaborated cognitive vulnerability-transactional stress theory. *Psychol. Bull.* **127**, 773–96 (2001).

151. D. S. Hasin, R. D. Goodwin, F. S. Stinson, B. F. Grant, Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Arch. Gen. Psychiatry*. **62**, 1097–106 (2005).

152. E. S. Paykel, T. Brugha, T. Fryers, Size and burden of depressive disorders in Europe. *Eur. Neuropsychopharmacol.* **15**, 411–23 (2005).

153. S. G. Kornstein, Gender differences in depression: Implications for treatment. *J. Clin. Psychiatry*. **58**, 12–18 (1997).

154. C. P. McLean, A. Asnaani, B. T. Litz, S. G. Hofmann, Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. *J. Psychiatr. Res.* **45**, 1027–35 (2011).

155. V. Simonds, V. Whiffen, Are gender differences in depression explained by gender differences in co-morbid anxiety? *J. Affect. Disord.* **77**, 197–202 (2003).

156. C. Kuehner, Gender differences in unipolar depression: an update of epidemiological findings and possible explanations. *Acta Psychiatr. Scand.* **108**, 163–74 (2003).

157. M. Piccinelli, G. Wilkinson, Gender differences in depression: Critical review. *Brit J Psychiat*. **177**, 486–492 (2000).

158. E. Young, A. Korszun, Sex, trauma, stress hormones and depression. *Mol. Psychiatry*. **15**, 23–8 (2010).

159. J. S. Hyde, A. H. Mezulis, L. Y. Abramson, The ABCs of depression: integrating affective, biological, and cognitive models to explain the emergence of the gender difference in depression. *Psychol. Rev.* **115**, 291–313 (2008).

160. C. N. Soares, B. Zitek, Reproductive hormone sensitivity and risk for depression across the female life cycle: a continuum of vulnerability? *J. Psychiatry Neurosci.* **33**, 331–43 (2008).

161. B. M. Kudielka, C. Kirschbaum, Sex differences in HPA axis responses to stress: a review. *Biol. Psychol.* **69**, 113–32 (2005).

162. L. R. Stroud, P. Salovey, E. S. Epel, Sex differences in stress responses: social rejection versus achievement stress. *Biol. Psychiatry.* **52**, 318–27 (2002).

163. H. Matsuzaka *et al.*, Gender differences in serum testosterone and cortisol in patients with major depressive disorder compared with controls. *Int. J. Psychiatry Med.* **46**, 203–21 (2013).

164. J. M. Goldstein, L. Holsen, R. Handa, S. Tobet, Fetal hormonal programming of sex differences in depression: linking women's mental health with sex differences in the brain across the lifespan. *Front. Neurosci.* **8**, 247 (2014).

165. K. Hinkelmann *et al.*, Sex differences of salivary cortisol secretion in patients with major depression. *Stress.* **15**, 105–9 (2012).

166. M. Owens *et al.*, Elevated morning cortisol is a stratified population-level biomarker for major depression in boys only with high depressive symptoms. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3638–43 (2014).

167. M. Danner, S. V. S. Kasl, J. L. J. Abramson, V. Vaccarino, Association between depression and elevated C-reactive protein. *Psychosom Med.* **65**, 347–356 (2003).

168. D. E. Ford, T. P. Erlinger, Depression and C-reactive protein in US adults: data from the Third National Health and Nutrition Examination Survey. *Arch. Intern. Med.* **164**, 1010–4 (2004).

169. S. Toker, A. Shirom, I. Shapira, S. Berliner, S. Melamed, The association between burnout, depression, anxiety, and inflammation biomarkers: C-reactive protein and fibrinogen in men and women. *J. Occup. Health Psychol.* **10**, 344–62 (2005).

170. T. Liukkonen *et al.*, The association between C-reactive protein levels and depression: Results from the northern Finland 1966 birth cohort study. *Biol. Psychiatry.* **60**, 825–30 (2006).

171. M. L. Vetter *et al.*, Gender differences in the relationship between symptoms of depression and high-sensitivity CRP. *Int. J. Obes. (Lond).* **37 Suppl 1**, S38–43 (2013).

172. A. A. Morris *et al.*, Association between depression and inflammation--differences by race and sex: the META-Health study. *Psychosom. Med.* **73**, 462–8.

173. J. McGrath, S. Saha, D. Chant, J. Welham, Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiol Rev.* **30**, 67–76 (2008).

174. M. DE Hert *et al.*, Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World Psychiatry.* **10**, 52–77 (2011).

175. World Health Organization (WHO), “The World Health Report 2001 - Mental Health: New Understanding, New Hope” (Geneva, 2001), (available at http://www.who.int/whr/2001/en/).

176. World Health Organization, “The global burden of disease: 2004 update” (Geneva, 2008).

177. W. Rössler, H. J. Salize, J. van Os, A. Riecher-Rössler, Size of burden of schizophrenia and psychotic disorders. *Eur. Neuropsychopharmacol.* **15**, 399–409 (2005).

178. M. Hambrecht, H. Häfner, Substance abuse and the onset of schizophrenia. *Biol. Psychiatry*. **40**, 1155–63 (1996).

179. L. Dixon, Dual diagnosis of substance abuse in schizophrenia: prevalence and impact on outcomes. *Schizophr. Res.* **35 Suppl**, S93–100 (1999).

180. S. Fazel, N. Långström, A. Hjern, M. Grann, P. Lichtenstein, Schizophrenia, substance abuse, and violent crime. *JAMA*. **301**, 2016–23 (2009).

181. J. de Leon, F. J. Diaz, A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors. *Schizophr. Res.* **76**, 135–57 (2005).

182. M. DE Hert, V. Schreurs, D. Vancampfort, R. VAN Winkel, Metabolic syndrome in people with schizophrenia: a review. *World Psychiatry*. **8**, 15–22 (2009).

183. R. Tandon *et al.*, Definition and description of schizophrenia in the DSM-5. *Schizophr. Res.* **150**, 3–10 (2013).

184. A. R. Yung *et al.*, Psychosis prediction: 12-month follow up of a high-risk (“prodromal”) group. *Schizophr. Res.* **60**, 21–32 (2003).

185. B. A. Palmer, V. S. Pankratz, J. M. Bostwick, The lifetime risk of suicide in schizophrenia: a reexamination. *Arch. Gen. Psychiatry*. **62**, 247–53 (2005).

186. R. Howard, P. V. Rabins, M. V. Seeman, D. V. Jeste, the I. Late-Onset, Late-Onset Schizophrenia and Very-Late-Onset Schizophrenia-Like Psychosis: An International Consensus. *Am. J. Psychiatry*. **157**, 172–8 (2014).

187. C. A. Ross, R. L. Margolis, S. A. J. Reading, M. Pletnikov, J. T. Coyle, Neurobiology of schizophrenia. *Neuron*. **52**, 139–53 (2006).

188. J. Geddes, N. Freemantle, P. Harrison, P. Bebbington, Atypical antipsychotics in the treatment of schizophrenia: systematic overview and meta-regression analysis. *BMJ*. **321**, 1371–1376 (2000).

189. S. Pilling *et al.*, Psychological treatments in schizophrenia: I. Meta-analysis of family intervention and cognitive behaviour therapy. *Psychol. Med.* **32**, 763–82 (2002).

190. G. Davies, J. Welham, D. Chant, E. F. Torrey, J. McGrath, A systematic review and meta-analysis of Northern Hemisphere season of birth studies in schizophrenia. *Schizophr. Bull.* **29**, 587–93 (2003).

191. J. van Os, S. Kapur, Schizophrenia. *Lancet*. **374**, 635–45 (2009).

192. I. C. Wright *et al.*, Meta-analysis of regional brain volumes in schizophrenia. *Am. J. Psychiatry*. **157**, 16–25 (2000).

193. P. J. Harrison, The neuropathology of schizophrenia: A critical review of the data and their interpretation. *Brain*. **122**, 593–624 (1999).

194. T. D. Cannon *et al.*, Regional gray matter, white matter, and cerebrospinal fluid distributions in schizophrenic patients, their siblings, and controls. *Arch. Gen. Psychiatry*. **55**, 1084–91 (1998).

195. B. Olabi *et al.*, Are there progressive brain changes in schizophrenia? A meta-analysis of structural magnetic resonance imaging studies. *Biol. Psychiatry*. **70**, 88–96 (2011).

196. B.-C. Ho, N. C. Andreasen, S. Ziebell, R. Pierson, V. Magnotta, Long-term antipsychotic treatment and brain volumes: a longitudinal study of first-episode schizophrenia. *Arch. Gen. Psychiatry*. **68**, 128–37 (2011).

197. W. Bunney, Evidence for a compromised dorsolateral prefrontal cortical parallel circuit in schizophrenia. *Brain Res. Rev.* **31**, 138–146 (2000).

198. S. Narayan *et al.*, Molecular profiles of schizophrenia in the CNS at different stages of illness. *Brain Res.* **1239**, 235–48 (2008).

199. M. P. Vawter *et al.*, Microarray analysis of gene expression in the prefrontal cortex in schizophrenia: a preliminary study. *Schizophr. Res.* **58**, 11–20 (2002).

200. K. Pennington *et al.*, Prominent synaptic and metabolic abnormalities revealed by proteomic analysis of the dorsolateral prefrontal cortex in schizophrenia and bipolar disorder. *Mol. Psychiatry*. **13**, 1102–17 (2008).

201. D. Martins-de-Souza *et al.*, Prefrontal cortex shotgun proteome analysis reveals altered calcium homeostasis and immune system imbalance in schizophrenia. *Eur. Arch. Psychiatry Clin. Neurosci.* **259**, 151–63 (2009).

202. D. Martins-de-Souza *et al.*, Proteomic analysis of dorsolateral prefrontal cortex indicates the involvement of cytoskeleton, oligodendrocyte, energy metabolism and new potential markers in schizophrenia. *J. Psychiatr. Res.* **43**, 978–86 (2009).

203. Y. Hakak *et al.*, Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4746–51 (2001).

204. D. Arion, T. Unger, D. A. Lewis, P. Levitt, K. Mirnics, Molecular evidence for increased expression of genes related to immune and chaperone function in the prefrontal cortex in schizophrenia. *Biol. Psychiatry*. **62**, 711–21 (2007).

205. J. K. Yao, J. A. Stanley, R. D. Reddy, M. S. Keshavan, J. W. Pettegrew, Correlations between peripheral polyunsaturated fatty acid content and in vivo membrane phospholipid metabolites. *Biol. Psychiatry*. **52**, 823–830 (2002).

206. G. P. Amminger *et al.*, Long-chain omega-3 fatty acids for indicated prevention of psychotic disorders: a randomized, placebo-controlled trial. *Arch. Gen. Psychiatry*. **67**, 146–54 (2010).

207. M. C. M. Ryan, P. Collins, J. H. Thakore, Impaired fasting glucose tolerance in first-episode, drug-naive patients with schizophrenia. *Am J Psychiatry*. **160**, 284–9 (2003).

208. D. Ben-Shachar *et al.*, Cerebral glucose utilization and platelet mitochondrial complex I activity in schizophrenia: A FDG-PET study. *Prog. Neuropsychopharmacol. Biol. Psychiatry*. **31**, 807–13 (2007).

209. P. C. Guest *et al.*, Altered levels of circulating insulin and other neuroendocrine hormones associated with the onset of schizophrenia. *Psychoneuroendocrinology*. **36**, 1092–6 (2011).

210. M. J. Green, S. L. Matheson, A. Shepherd, C. S. Weickert, V. J. Carr, Brain-derived neurotrophic factor levels in schizophrenia: a systematic review with meta-analysis. *Mol. Psychiatry*. **16**, 960–72 (2011).

211. B. J. Miller, P. Buckley, W. Seabolt, A. Mellor, B. Kirkpatrick, Meta-analysis of cytokine alterations in schizophrenia: clinical status and antipsychotic effects. *Biol. Psychiatry*. **70**, 663–71 (2011).

212. E. Schwarz *et al.*, Validation of a Blood-Based Laboratory Test to Aid in the Confirmation of a Diagnosis of Schizophrenia. *Biomark Insights*. **5**, 39–47 (2010).

213. L. W. Harris *et al.*, Comparison of peripheral and central schizophrenia biomarker profiles. *PLoS One*. **7**, e46368 (2012).

214. S. J. Glatt *et al.*, Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15533–8 (2005).

215. S. M. Purcell *et al.*, Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. **460**, 748–52 (2009).

216. S. Ripke *et al.*, Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. **511**, 421–427 (2014).

217. A. S. Bassett, S. W. Scherer, L. M. Brzustowicz, Copy number variations in schizophrenia: critical review and new perspectives on concepts of genetics and disease. *Am. J. Psychiatry*. **167**, 899–914 (2010).

218. T. L. Roth, F. D. Lubin, M. Sodhi, J. E. Kleinman, Epigenetic mechanisms in schizophrenia. *Biochim. Biophys. Acta*. **1790**, 869–77 (2009).

219. O. D. Howes, S. Kapur, The dopamine hypothesis of schizophrenia: version III--the final common pathway. *Schizophr. Bull.* **35**, 549–62 (2009).

220. G. Tsai, J. T. Coyle, Glutamatergic mechanisms in schizophrenia. *Annu. Rev. Pharmacol. Toxicol.* **42**, 165–79 (2002).

221. K. Nakazawa *et al.*, GABAergic interneuron origin of schizophrenia pathophysiology. *Neuropharmacology*. **62**, 1574–83 (2012).

222. A. E. Eggers, A serotonin hypothesis of schizophrenia. *Med. Hypotheses*. **80**, 791–4 (2013).

223. J. E. Lisman *et al.*, Circuit-based framework for understanding neurotransmitter and risk gene interactions in schizophrenia. *Trends Neurosci.* **31**, 234–42 (2008).

224. R. S. Smith, M. Maes, The macrophage-T-lymphocyte theory of schizophrenia: additional evidence. *Med. Hypotheses*. **45**, 135–41 (1995).

225. R. S. Smith, A comprehensive macrophage-T-lymphocyte theory of schizophrenia. *Med. Hypotheses.* **39**, 248–257 (1992).

226. Y. Watanabe, T. Someya, H. Nawa, Cytokine hypothesis of schizophrenia pathogenesis: Evidence from human studies and animal models. *Psychiatry Clin. Neurosci.* **64**, 217–230 (2010).

227. A. Monji, T. Kato, S. Kanba, Cytokines and schizophrenia: Microglia hypothesis of schizophrenia. *Psychiatry Clin. Neurosci.* **63**, 257–265 (2009).

228. M. J. Schwarz, N. Müller, M. Riedel, M. Ackenheil, The Th2-hypothesis of schizophrenia: a strategy to identify a subgroup of schizophrenia caused by immune mechanisms. *Med. Hypotheses*. **56**, 483–486 (2001).

229. U. Meyer, J. Feldon, B. K. Yee, A review of the fetal brain cytokine imbalance hypothesis of schizophrenia. *Schizophr. Bull.* **35**, 959–72 (2009).

230. D. R. Hanson, I. I. Gottesman, Theories of schizophrenia: a genetic-inflammatory-vascular synthesis. *BMC Med. Genet.* **6**, 7 (2005).

231. K. Bechter, Updating the mild encephalitis hypothesis of schizophrenia. *Prog. Neuropsychopharmacol. Biol. Psychiatry*. **42**, 71–91 (2013).

232. J. Steiner *et al.*, Increased prevalence of diverse N-methyl-D-aspartate glutamate receptor antibodies in patients with an initial diagnosis of schizophrenia: specific relevance of IgG NR1a antibodies for distinction from N-methyl-D-aspartate glutamate receptor encephalitis. *JAMA Psychiatry*. **70**, 271–8 (2013).

233. E. F. Walker, D. Diforio, Schizophrenia: a neural diathesis-stress model. *Psychol. Rev.* **104**, 667–85 (1997).

234. S. H. Fatemi, T. D. Folsom, The neurodevelopmental hypothesis of schizophrenia, revisited. *Schizophr. Bull.* **35**, 528–48 (2009).

235. J. Thome, P. Foley, P. Riederer, Neurotrophic factors and the maldevelopmental hypothesis of schizophrenic psychoses. *J. Neural Transm.* **105**, 85–100 (1998).

236. S. Ochoa, J. Usall, J. Cobo, X. Labad, J. Kulkarni, Gender differences in schizophrenia and first-episode psychosis: A comprehensive literature review. *Schizophr Res Treat.* **2012**, 916198 (2012).

237. A. Aleman, R. S. Kahn, J.-P. Selten, Sex differences in the risk of schizophrenia: evidence from meta-analysis. *Arch. Gen. Psychiatry*. **60**, 565–71 (2003).

238. A. Leung, P. Chue, Sex differences in schizophrenia, a review of the literature. *Acta Psychiatr Scand*. **101**, 3–38 (2000).

239. M. Hambrecht, K. Maurer, H. Häfner, N. Sartorius, Transnational stability of gender differences in schizophrenia? *Eur. Arch. Psychiatry Clin. Neurosci.* **242**, 6–12 (1992).

240. D. Castle, P. Sham, R. Murray, Differences in distribution of ages of onset in males and females with schizophrenia. *Schizophr. Res.* **33**, 179–183 (1998).

241. S. Beratis, J. Gabriel, S. Hoidas, Age at onset in subtypes of schizophrenic disorders. *Schizophr. Bull.* **20**, 287–96 (1994).

242. R. K. R. Salokangas, T. Honkonen, S. Saarinen, Women have later onset than men in schizophrenia—but only in its paranoid form. Results of the DSP project. *Eur. Psychiatry*. **18**, 274–281 (2003).

243. H. Häfner, Gender differences in schizophrenia. *Psychoneuroendocrinology*. **28**, 17–54 (2003).

244. A. Riecher-Rössler, H. Häfner, Gender aspects in schizophrenia: bridging the border between social and biological psychiatry. *Acta Psychiatr. Scand. Suppl.*, 58–62 (2000).

245. J. M. Goldstein, M. T. Tsuang, Gender and Schizophrenia: An Introduction and Synthesis of Findings. *Schizophr. Bull.* **16**, 179–183 (1990).

246. M. V Seeman, Gender differences in the prescribing of antipsychotic drugs. *Am. J. Psychiatry*. **161**, 1324–33 (2004).

247. P. Nopoulos, M. Flaum, N. C. Andreasen, Sex differences in brain morphology in schizophrenia. *Am. J. Psychiatry*. **154**, 1648–54 (1997).

248. I. C. Wright *et al.*, Meta-Analysis of Regional Brain Volumes in Schizophrenia. *Am. J. Psychiatry*. **157**, 16–25 (2000).

249. J. M. Goldstein *et al.*, Impact of normal sexual dimorphisms on sex differences in structural brain abnormalities in schizophrenia assessed by magnetic resonance imaging. *Arch. Gen. Psychiatry*. **59**, 154–64 (2002).

250. R. E. Gur *et al.*, A sexually dimorphic ratio of orbitofrontal to amygdala volume is altered in schizophrenia. *Biol. Psychiatry*. **55**, 512–7 (2004).

251. A. Mendrek, A. Mancini-Marië, C. Fahim, E. Stip, Sex differences in the cerebral function associated with processing of aversive stimuli by schizophrenia patients. *Aust. N. Z. J. Psychiatry*. **41**, 136–41 (2007).

252. J. A. Jiménez, A. Mancini-Marïe, N. Lakis, M. Rinaldi, A. Mendrek, Disturbed sexual dimorphism of brain activation during mental rotation in schizophrenia. *Schizophr. Res.* **122**, 53–62 (2010).

253. J. M. Goldstein *et al.*, Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. *Biol. Psychiatry*. **61**, 935–45 (2007).

254. R. E. Gur, B. I. Turetsky, W. B. Bilker, R. C. Gur, Reduced gray matter volume in schizophrenia. *Arch. Gen. Psychiatry*. **56**, 905–11 (1999).

255. R. E. Gur *et al.*, Reduced dorsal and orbital prefrontal gray matter volumes in schizophrenia. *Arch. Gen. Psychiatry*. **57**, 761–8 (2000).

256. T. E. Schlaepfer *et al.*, Decreased regional cortical gray matter volume in schizophrenia. *Am. J. Psychiatry*. **151**, 842–8 (1994).

257. P. E. Cowell, D. J. Kostianovsky, R. C. Gur, B. I. Turetsky, R. E. Gur, Sex differences in neuroanatomical and clinical correlations in schizophrenia. *Am. J. Psychiatry*. **153**, 799–805 (1996).

258. T. J. Crow, Cerebral asymmetry and the lateralization of language: core deficits in schizophrenia as pointers to the gene. *Curr. Opin. Psychiatry*. **17**, 97–106 (2004).

259. R. J. Hennessy *et al.*, 3D morphometrics of craniofacial dysmorphology reveals sex-specific asymmetries in schizophrenia. *Schizophr. Res.* **67**, 261–8 (2004).

260. M. V Seeman, M. Lang, The role of estrogens in schizophrenia gender differences. *Schizophr. Bull.* **16**, 185–94 (1990).

261. M. V Seeman, The role of estrogen in schizophrenia. *J. Psychiatry Neurosci.* **21**, 123–7 (1996).

262. N. Bergemann, P. Parzer, B. Runnebaum, F. Resch, C. Mundt, Estrogen, menstrual cycle phases, and psychopathology in women suffering from schizophrenia. *Psychol. Med.* **37**, 1427–36 (2007).

263. A. Riecher-Rössler, H. Häfner, M. Stumbaum, K. Maurer, R. Schmidt, Can estradiol modulate schizophrenic symptomatology? *Schizophr. Bull.* **20**, 203–14 (1994).

264. D. La Torre, A. Falorni, Pharmacological causes of hyperprolactinemia. *Ther. Clin. Risk Manag.* **3**, 929–51 (2007).

265. U. Halbreich, B. J. Kinon, J. A. Gilmore, L. S. Kahn, Elevated prolactin levels in patients with schizophrenia: mechanisms and related adverse effects. *Psychoneuroendocrinology*. **28**, 53–67 (2003).

266. A. Riecher-Rössler, H. Häfner, A. Dütsch-Strobel, M. Stumbaum, Gonadal function and its influence on psychopathology. *Arch. Womens. Ment. Health*. **1**, 15–26 (1998).

267. N. Bergemann *et al.*, Plasma concentrations of estradiol in women suffering from schizophrenia treated with conventional versus atypical antipsychotics. *Schizophr. Res.* **73**, 357–66 (2005).

268. E. Schwarz *et al.*, Identification of a biological signature for schizophrenia in serum. *Mol. Psychiatry*. **17**, 1–9 (2011).

269. S. Akhondzadeh *et al.*, Correlation between testosterone, gonadotropins and prolactin and severity of negative symptoms in male patients with chronic schizophrenia. *Schizophr. Res.* **84**, 405–10 (2006).

270. Y.-H. Ko *et al.*, Association between serum testosterone levels and the severity of negative symptoms in male patients with chronic schizophrenia. *Psychoneuroendocrinology*. **32**, 385–91 (2007).

271. J. M. Goldstein, S. Cherkerzian, M. T. Tsuang, T. L. Petryshen, Sex differences in the genetic risk for schizophrenia: history of the evidence for sex-specific and sex-dependent effects. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **162B**, 698–710 (2013).

272. S. C. Godar, M. Bortolato, Gene-sex interactions in schizophrenia: focus on dopamine neurotransmission. *Front. Behav. Neurosci.* **8**, 71 (2014).

273. J. Salem, The role of gender differences in the reduction of etiologic heterogeneity in schizophrenia. *Clin. Psychol. Rev.* **18**, 795–819 (1998).

274. D. J. Castle, R. M. Murray, The neurodevelopmental basis of sex differences in schizophrenia. *Psychol. Med.* **21**, 565–75 (1991).

275. R. Murray, E. O'Callaghan, D. Castle, S. Lewis, A neurodevelopmental approach to the classification of schizophrenia. *Schizophr. Bull.* **18**, 319–32 (1992).

276. J. M. Goldstein, S. L. Santangelo, J. C. Simpson, M. T. Tsuang, The role of gender in identifying subtypes of schizophrenia: a latent class analytic approach. *Schizophr. Bull.* **16**, 263–75 (1990).

277. A. Mendrek, Reversal of normal cerebral sexual dimorphism in schizophrenia: evidence and speculations. *Med Hypotheses*. **69**, 896–902 (2007).

278. E. Schwarz *et al.*, Identification of a biological signature for schizophrenia in serum. *Mol Psychiatry*. **17**, 494–502 (2012).

279. F. Haenisch *et al.*, Multiplex immunoassay analysis of plasma shows prominent upregulation of growth factor activity pathways linked to GSK3β signaling in bipolar patients. *J. Affect. Disord.* **156**, 139–43 (2014).

280. J. M. Ramsey *et al.*, Identification of an age-dependent biomarker signature in children and adolescents with autism spectrum disorders. *Mol. Autism*. **4**, 27 (2013).

281. J. M. Ramsey *et al.*, Distinct molecular phenotypes in male and female schizophrenia patients. *PLoS One*. **8**, e78729 (2013).

282. N. Delaleu, H. Immervoll, J. Cornelius, R. Jonsson, Biomarker profiles in serum and saliva of experimental Sjögren's syndrome: associations with specific autoimmune manifestations. *Arthritis Res Ther*. **10**, R22 (2008).

283. T. M. Umstead *et al.*, The kinetics of cardiopulmonary bypass: a dual-platform proteomics study of plasma biomarkers in pediatric patients undergoing cardiopulmonary bypass. *Artif Organs*. **36**, E1–20 (2012).

284. G. P. Bertenshaw *et al.*, Multianalyte profiling of serum antigens and autoimmune and infectious disease molecules to identify biomarkers dysregulated in epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev*. **17**, 2872–81 (2008).

285. R Core Team, R: A Language and Environment for Statistical Computing (2014), (available at http://www.r-project.org/).

286. D. D. Dalma-Weiszhausz, J. Warrington, E. Y. Tanimoto, C. G. Miyada, The Affymetrix GeneChip platform: an overview. *Methods Enzymol.* **410**, 3–28 (2006).

287. Affymetrix, “Technical Note: Design and Performance of the GeneChip Human Genome U133 Plus 2.0 and Human Genome U133A 2.0 Arrays.”

288. Affymetrix, “Data Sheet: GeneChip Human Genome U133 Arrays.”

289. D. Greenbaum, C. Colangelo, K. Williams, M. Gerstein, Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* **4**, 117 (2003).

290. B. Schwanhäusser *et al.*, Global quantification of mammalian gene expression control. *Nature*. **473**, 337–42 (2011).

291. P. R. Maycox *et al.*, Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. *Mol. Psychiatry*. **14**, 1083–94 (2009).

292. K. Iwamoto, M. Bundo, T. Kato, Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis. *Hum. Mol. Genet.* **14**, 241–53 (2005).

293. E. F. Torrey, M. Webster, M. Knable, N. Johnston, R. H. Yolken, The Stanley Foundation brain collection and neuropathology consortium. *Schizophr. Res.* **44**, 151–5 (2000).

294. R. C. Gentleman *et al.*, Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

295. A. Kauffmann, R. Gentleman, W. Huber, arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. *Bioinformatics*. **25**, 415–6 (2009).

296. L. Gautier, L. Cope, B. M. Bolstad, R. A. Irizarry, affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. **20**, 307–15 (2004).

297. C. L. Wilson, C. J. Miller, Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics*. **21**, 3683–5 (2005).

298. M. N. McCall, B. M. Bolstad, R. a Irizarry, Frozen robust multiarray analysis (fRMA). *Biostatistics*. **11**, 242–53 (2010).

299. B. M. Bolstad, thesis, University of California, Berkeley (2004).

300. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43** (2015), doi:10.1093/nar/gkv007.

301. G. K. Smyth, Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).

302. M. E. Ritchie *et al.*, Empirical array quality weights in the analysis of microarray data. *BMC Bioinformatics*. **7**, 261 (2006).

303. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. **8**, 118–27 (2007).

304. J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, J. D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. **28**, 882–3 (2012).

305. P. Filzmoser, R. Maronna, M. Werner, Outlier identification in high dimensions. *Comput. Stat. Data Anal.* **52**, 1694–1711 (2008).

306. R. A. Maronna, R. H. Zamar, Robust Estimates of Location and Dispersion for High-Dimensional Datasets. *Technometrics*. **44**, 307–317 (2002).

307. J. S. Hardin, D. M. Rocke, The Distribution of Robust Distances. *J. Comput. Graph. Stat.* **14**, 928–946 (2005).

308. J. Wang *et al.*, robust: Robust Library (2013), (available at http://cran.r-project.org/package=robust).

309. P. Filzmoser, M. Gschwandtner, mvoutlier: Multivariate outlier detection based on robust methods (2014), (available at http://cran.r-project.org/package=mvoutlier).

310. J. Shlens, A Tutorial on Principal Component Analysis. *arXiv* (2009).

311. F. Husson, J. Josse, S. Le, J. Mazet, FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R (2014), (available at http://cran.r-project.org/package=FactoMineR).

312. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, NY, ed. 2, 2009).

313. G. A. F. Seber, A. J. Lee, *Linear Regression Analysis* (John Wiley & Sons, Inc., Hoboken, NJ, ed. 2, 2003).

314. M. A. Poole, P. N. O'Farrell, The Assumptions of the Linear Regression Model. *Trans. Inst. Br. Geogr.* **52**, 145–158 (1971).

315. J. Fox, *Regression Diagnostics* (SAGE Publications, Inc, Thousand Oaks, CA, 1991; http://dx.doi.org/10.4135/9781412985604).

316. J. Fox, S. Weisberg, *An {R} Companion to Applied Regression* (Sage, Thousand Oaks, CA, Second., 2011; http://socserv.socsci.mcmaster.ca/jfox/Books/Companion).

317. P. Rousseeuw *et al.*, robustbase: Basic Robust Statistics (2014), (available at http://cran.r-project.org/package=robustbase).

318. R. Anderson, Ed., *Modern Methods for Robust Regression* (SAGE Publications, Inc., Thousand Oaks, CA, 2008; http://dx.doi.org/10.4135/9781412985109).

319. V. J. Yohai, High Breakdown-Point and High Efficiency Robust Estimates for Regression. *Ann. Stat.* **15**, 642–656 (1987).

320. J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, R Core Team, nlme: Linear and Nonlinear Mixed Effects Models (2013).

321. S. Menard, *Applied Logistic Regression Analysis* (SAGE Publications, Inc., Thousand Oaks, CA, ed. 2, 2002).

322. G. P. Quinn, M. J. Keough, *Experimental Design and Data Analysis for Biologists* (Cambridge University Press, Cambridge, UK, 2002).

323. J. C. Stoltzfus, Logistic regression: a brief primer. *Acad. Emerg. Med.* **18**, 1099–104 (2011).

324. J. Lye, J. Hirschberg, Tests of Inference for Dummy Variables in Regressions with Logarithmic Transformed Dependent Variables (2002).

325. G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R* (Springer, New York, NY, ed. 1, 2013).

326. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Stat Methodol*. **57**, 289–300 (1995).

327. L. Breiman, Random Forests. *Mach. Learn.* **45**, 5–32 (2001).

328. L. Breiman, "Manual on Setting Up, Using, and Understanding Random Forests V3.1" (2002), (available at http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1.pdf).

329. A. Liaw, M. Wiener, Classification and Regression by randomForest. *R News*. **2**, 18–22 (2002).

330. R. Kohavi, in *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2* (Morgan Kaufmann Publishers Inc., San Francisco, CA, 1995; http://dl.acm.org/citation.cfm?id=1643031.1643047), pp. 1137–1143.

331. T. Fawcett, An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874 (2006).

332. L. Yan, R. Dodier, M. C. Mozer, R. Wolniewicz, in *Proceedings of the 20th International Conference on Machine Learning (ICML-2003)*, T. Fawcett, N. Mishra, Eds. (AAAI Press, Menlo Park, CA, 2003).

333. T. Sing, O. Sander, N. Beerenwinkel, T. Lengauer, ROCR: visualizing classifier performance in R. *Bioinformatics*. **21**, 3940–1 (2005).

334. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–9 (2000).

335. M. Carlson, UniProt.ws: R Interface to UniProt Web Services (2015).

336. S. Durinck *et al.*, BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*. **21**, 3439–40 (2005).

337. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–91 (2009).

338. H. Fröhlich, N. Speer, A. Poustka, T. Beissbarth, GOSim--an R-package for computation of information theoretic GO similarities between terms and gene products. *BMC Bioinformatics*. **8**, 166 (2007).

339. S. Falcon, R. Gentleman, Using GOstats to test gene lists for GO term association. *Bioinformatics*. **23**, 257–8 (2007).

340. T. T. Rivet, J. L. Matson, Review of gender differences in core symptomatology in autism spectrum disorders. *Res. Autism Spectr. Disord.* **5**, 957–976 (2011).

341. J. M. Ramsey *et al.*, Molecular sex differences in human serum. *PLoS One*. **7**, e51504 (2012).

342. B. W. J. H. Penninx *et al.*, The Netherlands Study of Depression and Anxiety (NESDA): Rationale, objectives and methods. *Int J Methods Psychiatr Res*. **17**, 121–140 (2008).

343. C. L. Craig *et al.*, International physical activity questionnaire: 12-country reliability and validity. *Med. Sci. Sports Exerc.* **35**, 1381–95 (2003).

344. A. J. Rush, C. M. Gullion, M. R. Basco, R. B. Jarrett, M. H. Trivedi, The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychol. Med.* **26**, 477–86 (1996).

345. A. T. Beck, N. Epstein, G. Brown, R. A. Steer, An inventory for measuring clinical anxiety: psychometric properties. *J. Consult. Clin. Psychol.* **56**, 893–7 (1988).

346. P. C. Guest, S. Bahn, Eds., *Biomarkers of Neurological and Psychiatric Disease* (Elsevier Inc., San Diego, CA, ed. 1, 2011).

347. P. C. Hannaford *et al.*, Mortality among contraceptive pill users: cohort evidence from Royal College of General Practitioners' Oral Contraception Study. *BMJ*. **340**, c927 (2010).

348. K. Sherif, Benefits and risks of oral contraceptives. *Am. J. Obstet. Gynecol.* **180**, S343–S348 (1999).

349. J. A. Cornish *et al.*, The risk of oral contraceptives in the etiology of inflammatory bowel disease: a meta-analysis. *Am. J. Gastroenterol.* **103**, 2394–400 (2008).

350. K. M. Keyes *et al.*, Association of hormonal contraceptive use with reduced levels of depressive symptoms: a national study of sexually active women in the United States. *Am. J. Epidemiol.* **178**, 1378–88 (2013).

351. D. G. M. C. Rosano, C. Vitale, G. Marazzi, M. Volterrani, Menopause and cardiovascular disease: the evidence (2009) (available at http://informahealthcare.com/doi/abs/10.1080/13697130601114917).

352. B. L. Riggs, Pathogenesis of osteoporosis. *Am. J. Obstet. Gynecol.* **156**, 1342–1346 (1987).

353. M. C. Carr, The Emergence of the Metabolic Syndrome with Menopause. *J. Clin. Endocrinol. Metab.* **88**, 2404–2411 (2003).

354. M. X. Tang *et al.*, Effect of oestrogen during menopause on risk and age at onset of Alzheimer's disease. *Lancet.* **348**, 429–32 (1996).

355. J. A. P. Silva, Sex Hormones and Glucocorticoids: Interactions with the Immune System. *Ann. N. Y. Acad. Sci.* **876**, 102–118 (1999).

356. A. H. Miller, V. Maletic, C. L. Raison, Inflammation and its discontents: the role of cytokines in the pathophysiology of major depression. *Biol. Psychiatry.* **65**, 732–41 (2009).

357. G. Venkatasubramanian *et al.*, Insulin and insulin-like growth factor-1 abnormalities in antipsychotic-naive schizophrenia. *Am. J. Psychiatry.* **164**, 1557–60 (2007).

358. P. Videbech, PET measurements of brain glucose metabolism and blood flow in major depressive disorder: a critical review. *Acta Psychiatr. Scand.* **101**, 11–20 (2000).

359. G. Venkatasubramanian *et al.*, Insulin and insulin-like growth factor-1 abnormalities in antipsychotic-naive schizophrenia. *Am. J. Psychiatry.* **164**, 1557–60 (2007).

360. B. A. Hoshaw, J. E. Malberg, I. Lucki, Central administration of IGF-I and BDNF leads to long-lasting antidepressant-like effects. *Brain Res.* **1037**, 204–8 (2005).

361. F. Liu *et al.*, The link between angiotensin II-mediated anxiety and mood disorders with NADPH oxidase-induced oxidative stress. *Int. J. Physiol. Pathophysiol. Pharmacol.* **4**, 28–35 (2012).

362. O. P. Almeida *et al.*, Angiogenesis inhibition and depression in older men. *J. Psychiatry Neurosci.* **39**, 200–5 (2014).

363. R. Lopes, R. Soares, M. Figueiredo-Braga, R. Coelho, Schizophrenia and cancer: is angiogenesis a missed link? *Life Sci.* **97**, 91–5 (2014).

364. S. J. James *et al.*, Metabolic biomarkers of increased oxidative stress and impaired methylation capacity in children with autism. *Am J Clin Nutr.* **80**, 1611–1617 (2004).

365. A. B. Berenson, C. R. Breitkopf, Z. H. Wu, Reproductive correlates of depressive symptoms among low-income minority women. *Obstet. Gynecol.* **102**, 1310–7 (2003).

366. L. S. Callegari, X. Zhao, K. M. Nelson, S. Borrero, Contraceptive adherence among women Veterans with mental illness and substance use disorder. *Contraception.* **91**, 386–92 (2015).

367. M. J. Kurver, C. L. van der Wijden, J. Burgers, "Summary of the Dutch College of General Practitioners' Practice Guideline 'Contraception'" (2012), (available at http://www.ntvg.nl/).

368. A. van Hylckama Vlieg, F. M. Helmerhorst, J. P. Vandenbroucke, C. J. M. Doggen, F. R. Rosendaal, The venous thrombotic risk of oral contraceptives, effects of oestrogen dose and progestogen type: results of the MEGA case-control study. *BMJ.* **339**, b2921 (2009).

369. H. D. Schmidt, R. C. Shelton, R. S. Duman, Functional biomarkers of depression: diagnosis, treatment, and pathophysiology. *Neuropsychopharmacology.* **36**, 2375–94 (2011).

370. H. U. Wittchen, M. Höfler, W. Meister, Prevalence and recognition of depressive syndromes in German primary care settings: poorly recognized and treated? *Int. Clin. Psychopharmacol.* **16**, 121–35 (2001).

371. X. Ding *et al.*, The potential biomarker panels for identification of Major Depressive Disorder (MDD) patients with and without early life stress (ELS) by metabonomic analysis. *PLoS One*. **9**, e97479 (2014).

372. E. E. Redei *et al.*, Blood transcriptomic biomarkers in adult primary care patients with major depressive disorder undergoing cognitive behavioral therapy. *Transl. Psychiatry*. **4**, e442 (2014).

373. M. Bot *et al.*, Serum proteomic profiling of major depressive disorder. *Transl. Psychiatry*. **5**, e599 (2015).

374. A. J. Fyer, M. M. Weissman, Genetic linkage study of panic: clinical methodology and description of pedigrees. *Am. J. Med. Genet.* **88**, 173–81 (1999).

375. N. Vogelzangs *et al.*, Association of depressive disorders, depression characteristics and antidepressant medication with inflammation. *Transl. Psychiatry*. **2**, e79 (2012).

376. G. P. Chrousos, Stress and sex versus immunity and inflammation. *Sci. Signal.* **3**, pe36 (2010).

377. C. H. Bourke, C. S. Harrell, G. N. Neigh, Stress-induced sex differences: adaptations mediated by the glucocorticoid receptor. *Horm. Behav.* **62**, 210–8 (2012).

378. J. C. Stewart, K. L. Rand, M. F. Muldoon, T. W. Kamarck, A prospective evaluation of the directionality of the depression-inflammation relationship. *Brain. Behav. Immun.* **23**, 936–44 (2009).

379. S. L. Klein, The effects of hormones on sex differences in infection: from genes to behavior. *Neurosci. Biobehav. Rev.* **24**, 627–638 (2000).

380. D. Fairweather, S. Frisancho-Kiss, N. R. Rose, Sex differences in autoimmune disease from a pathological perspective. *Am. J. Pathol.* **173**, 600–9 (2008).

381. H. Steeb *et al.*, Serum proteomic analysis identifies sex-specific differences in lipid metabolism and inflammation profiles in adults diagnosed with Asperger syndrome. *Mol. Autism*. **5**, 4 (2014).

382. E. Kajantie, D. I. W. Phillips, The effects of sex and hormonal status on the physiological response to acute psychosocial stress. *Psychoneuroendocrinology*. **31**, 151–78 (2006).

383. A. Kerchner, W. Lester, S. P. Stuart, A. Dokras, Risk of depression and other mental health disorders in women with polycystic ovary syndrome: a longitudinal study. *Fertil. Steril.* **91**, 207–12 (2009).

384. C. L. Weiner, M. Primeau, D. A. Ehrmann, Androgens and mood dysfunction in women: comparison of women with polycystic ovarian syndrome to healthy controls. *Psychosom. Med.* **66**, 356–62 (2004).

385. U. D. Rohr, The impact of testosterone imbalance on depression and women's health. *Maturitas*. **41 Suppl 1**, S25–46 (2002).

386. E. A. Young, A. R. Midgley, N. E. Carlson, M. B. Brown, Alteration in the hypothalamic-pituitary-ovarian axis in depressed women. *Arch. Gen. Psychiatry*. **57**, 1157–62 (2000).

387. B. L. Harlow, L. A. Wise, M. W. Otto, C. N. Soares, L. S. Cohen, Depression and its influence on reproductive endocrine and menstrual cycle markers associated with perimenopause: the Harvard Study of Moods and Cycles. *Arch. Gen. Psychiatry*. **60**, 29–36 (2003).

388. S. Potvin *et al.*, Inflammatory cytokine alterations in schizophrenia: a systematic quantitative review. *Biol. Psychiatry*. **63**, 801–8 (2008).

389. E. Schwarz *et al.*, Identification of Subgroups of Schizophrenia Patients With Changes in Either Immune or Growth Factor and Hormonal Pathways. *Schizophr. Bull.* **40**, 787–95 (2013).

390. M. K. Al Kindi, F. S. Al Essry, F. S. Al Essry, W.-A. S. Mula-Abed, Validity of serum testosterone, free androgen index, and calculated free testosterone in women with suspected hyperandrogenism. *Oman Med. J.* **27**, 471–4 (2012).

391. D. M. Corey, W. P. Dunlap, J. Michael, Averaging correlations: Expected values and bias in combined Pearson rs and Fisher's z transformations. *J Gen Psychol*. **125**, 245–61 (1998).

392. S. Benson *et al.*, Disturbed stress responses in women with polycystic ovary syndrome. *Psychoneuroendocrinology*. **34**, 727–35 (2009).

393. C. Invitti, M. De Martin, G. Delitala, J. D. Veldhuis, F. Cavagnini, Altered morning and nighttime pulsatile corticotropin and cortisol release in polycystic ovary syndrome. *Metabolism.* **47**, 143–8 (1998).

394. A. J. Morales *et al.*, Insulin, somatotropic, and luteinizing hormone axes in lean and obese women with polycystic ovary syndrome: common and distinct features. *J. Clin. Endocrinol. Metab.* **81**, 2854–64 (1996).

395. M. Salehi, R. Bravo-Vera, A. Sheikh, A. Gouller, L. Poretsky, Pathogenesis of polycystic ovary syndrome: what is the role of obesity? *Metabolism*. **53**, 358–376 (2004).

396. D. A. Dumesic, D. H. Abbott, V. Padmanabhan, Polycystic ovary syndrome and its developmental origins. *Rev. Endocr. Metab. Disord.* **8**, 127–41 (2007).

397. S. Franks, S. L. Berga, Does PCOS have developmental origins? *Fertil. Steril.* **97**, 2–6 (2012).

398. D. L. Kelly, R. R. Conley, Sexuality and schizophrenia: a review. *Schizophr. Bull.* **30**, 767–79 (2004).

399. M. Di Nicola *et al.*, Serum and gene expression profile of cytokines in first-episode psychosis. *Brain. Behav. Immun.* **31**, 90–5 (2013).

400. S. Fulzele, A. Pillai, Decreased VEGF mRNA expression in the dorsolateral prefrontal cortex of schizophrenia subjects. *Schizophr. Res.* **115**, 372–3 (2009).

401. T. Yasuhara *et al.*, Neurorescue effects of VEGF on a rat model of Parkinson's disease. *Brain Res.* **1053**, 10–8 (2005).

402. S. Takano *et al.*, Concentration of Vascular Endothelial Growth Factor in the Serum and Tumor Tissue of Brain Tumor Patients. *Cancer Res.* **56**, 2185–2190 (1996).

403. B. Elfving, P. H. Plougmann, G. Wegener, Differential brain, but not serum VEGF levels in a genetic rat model of depression. *Neurosci. Lett.* **474**, 13–6 (2010).

404. N. Husain, S. Awasthi, M. Haris, R. K. Gupta, M. Husain, Vascular endothelial growth factor as a marker of disease activity in neurotuberculosis. *J. Infect.* **56**, 114–9 (2008).

405. K. Fabel *et al.*, VEGF is necessary for exercise-induced adult hippocampal neurogenesis. *Eur. J. Neurosci.* **18**, 2803–2812 (2003).

406. T. A. Jenkins, F. A. Mendelsohn, S. Y. Chai, Angiotensin-converting enzyme modulates dopamine turnover in the striatum. *J. Neurochem.* **68**, 1304–11 (1997).

407. R. J. Banks, L. Mozley, C. T. Dourish, The angiotensin converting enzyme inhibitors captopril and enalapril inhibit apomorphine-induced oral stereotypy in the rat. *Neuroscience*. **58**, 799–805 (1994).

408. M. van den Buuse, T. W. Zheng, L. L. Walker, D. a Denton, Angiotensin-converting enzyme (ACE) interacts with dopaminergic mechanisms in the brain to modulate prepulse inhibition in mice. *Neurosci. Lett.* **380**, 6–11 (2005).

409. K. A. Reardon, F. A. Mendelsohn, S. Y. Chai, M. K. Horne, The angiotensin converting enzyme (ACE) inhibitor, perindopril, modifies the clinical features of Parkinson's disease. *Aust. N. Z. J. Med.* **30**, 48–53 (2000).

410. K. Wahlbeck, A. Ahokas, H. Nikkilä, K. Miettinen, R. Rimón, Cerebrospinal fluid angiotensin-converting enzyme (ACE) correlates with length of illness in schizophrenia. *Schizophr Res*. **41**, 335–40 (2000).

411. M. Rothermundt, V. Arolt, T. A. Bayer, Review of immunological and immunopathological findings in schizophrenia. *Brain. Behav. Immun.* **15**, 319–39 (2001).

412. A. T. Vella, S. Dow, T. A. Potter, J. Kappler, P. Marrack, Cytokine-induced survival of activated T cells in vitro and in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3810–3815 (1998).

413. L. P. Perera, C. K. Goldman, T. A. Waldmann, IL-15 induces the expression of chemokines and their receptors in T lymphocytes. *J. Immunol.* **162**, 2606–12 (1999).

414. S. Columba-Cabezas *et al.*, Induction of macrophage-derived chemokine/CCL22 expression in experimental autoimmune encephalomyelitis and cultured microglia: implications for disease regulation. *J. Neuroimmunol.* **130**, 10–21 (2002).

415. V. Mondelli, O. Howes, Inflammation: its role in schizophrenia and the potential anti-inflammatory effects of antipsychotics. *Psychopharmacology (Berl).* **231**, 317–8 (2014).

416. A. K. Sárvári, Z. Veréb, I. P. Uray, L. Fésüs, Z. Balajthy, Atypical antipsychotics induce both proinflammatory and adipogenic gene expression in human adipocytes in vitro. *Biochem. Biophys. Res. Commun.* **450**, 1383–9 (2014).

417. J. M. Meyer *et al.*, Inflammatory markers in schizophrenia: comparing antipsychotic effects in phase 1 of the clinical antipsychotic trials of intervention effectiveness study. *Biol. Psychiatry*. **66**, 1013–22 (2009).

418. H. Sugino, T. Futamura, Y. Mitsumoto, K. Maeda, Y. Marunaka, Atypical antipsychotics suppress production of proinflammatory cytokines and up-regulate interleukin-10 in lipopolysaccharide-treated mice. *Prog. Neuropsychopharmacol. Biol. Psychiatry*. **33**, 303–7 (2009).

419. T. A. Kato *et al.*, Anti-Inflammatory properties of antipsychotics via microglia modulations: are antipsychotics a “fire extinguisher” in the brain of schizophrenia? *Mini Rev. Med. Chem.* **11**, 565–74 (2011).

420. R.-R. Wu *et al.*, Metformin for treatment of antipsychotic-induced amenorrhea and weight gain in women with first-episode schizophrenia: A double-blind, randomized, placebo-controlled study. *Am. J. Psychiatry*. **169**, 813–821 (2012).

421. W. Laan *et al.*, Non-steroidal anti-inflammatory drugs and the risk of psychosis. *Eur. Neuropsychopharmacol.* **17**, 309–11 (2007).

422. W. Laan *et al.*, Glucocorticosteroids associated with a decreased risk of psychosis. *J. Clin. Psychopharmacol.* **29**, 288–90 (2009).

423. I. E. Sommer *et al.*, Efficacy of anti-inflammatory agents to improve symptoms in patients with schizophrenia: an update. *Schizophr. Bull.* **40**, 181–91 (2014).

424. S. Narayan, S. R. Head, T. J. Gilmartin, B. Dean, E. A. Thomas, Evidence for disruption of sphingolipid metabolism in schizophrenia. *J. Neurosci. Res.* **87**, 278–88 (2009).

425. T. Sugai *et al.*, Prefrontal abnormality of schizophrenia revealed by DNA microarray: impact on glial and neurotrophic gene expression. *Ann. N. Y. Acad. Sci.* **1025**, 84–91 (2004).

426. D. Tkachev *et al.*, Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *Lancet*. **362**, 798–805 (2003).

427. K. Mirnics, F. A. Middleton, A. Marquez, D. A. Lewis, P. Levitt, Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron*. **28**, 53–67 (2000).

428. T. Hashimoto *et al.*, Alterations in GABA-related transcriptome in the dorsolateral prefrontal cortex of subjects with schizophrenia. *Mol. Psychiatry*. **13**, 147–61 (2008).

429. J. Pérez-Santiago *et al.*, A combined analysis of microarray gene expression studies of the human prefrontal cortex identifies genes implicated in schizophrenia. *J. Psychiatr. Res.* **46**, 1464–74 (2012).

430. M. Mistry, J. Gillis, P. Pavlidis, Genome-wide expression profiling of schizophrenia using a large combined cohort. *Mol. Psychiatry*. **18**, 1–11 (2012).

431. M. L. Collaer, M. Hines, Human behavioral sex differences: a role for gonadal hormones during early development? *Psychol. Bull.* **118**, 55–107 (1995).

432. J. M. Goldstein, Normal Sexual Dimorphism of the Adult Human Brain Assessed by In Vivo Magnetic Resonance Imaging. *Cereb. Cortex*. **11**, 490–497 (2001).

433. O. Speck *et al.*, Gender differences in the functional organization of the brain for working memory. *Neuroreport*. **11**, 2581–5 (2000).

434. J. M. Goldstein *et al.*, Sex differences in prefrontal cortical brain activity during fMRI of auditory verbal working memory. *Neuropsychology*. **19**, 509–19 (2005).

435. R. M. Shansky *et al.*, Estrogen mediates sex differences in stress-induced prefrontal cortex dysfunction. *Mol. Psychiatry*. **9**, 531–8 (2004).

436. H. Xu *et al.*, Sex-biased methylome and transcriptome in human prefrontal cortex. *Hum. Mol. Genet.* **23**, 1260–70 (2014).

437. C. S. Weickert *et al.*, Transcriptome analysis of male-female differences in prefrontal cortical development. *Mol. Psychiatry*. **14**, 558–61 (2009).

438. M. P. Vawter *et al.*, Gender-specific gene expression in post-mortem human brain: localization to sex chromosomes. *Neuropsychopharmacology*. **29**, 373–84 (2004).

439. J. M. Goldstein *et al.*, Are there sex differences in neuropsychological functions among patients with schizophrenia? *Am. J. Psychiatry*. **155**, 1358–64 (1998).

440. D. Roesch-Ely *et al.*, Do sex differences affect prefrontal cortex associated cognition in schizophrenia? *Schizophr. Res.* **107**, 255–61 (2009).

441. M. Petrides, D. N. Pandya, Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* **7**, e1000170 (2009).

442. G. Liu *et al.*, NetAffx: Affymetrix probesets and annotations. *Nucleic Acids Res.* **31**, 82–6 (2003).

443. G. K. Smyth, in *Bioinformatics and Computational Biology Solutions using R and Bioconductor.*, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber, Eds. (Springer, New York, NY, 2005), pp. 397–420.

444. W. R. Perlman, M. J. Webster, J. E. Kleinman, C. S. Weickert, Reduced glucocorticoid and estrogen receptor alpha messenger ribonucleic acid levels in the amygdala of patients with major mental illness. *Biol. Psychiatry*. **56**, 844–52 (2004).

445. M. J. Webster, M. B. Knable, J. O'Grady, J. Orthmann, C. S. Weickert, Regional specificity of brain glucocorticoid receptor mRNA alterations in subjects with schizophrenia and mood disorders. *Mol. Psychiatry*. **7**, 985–94, 924 (2002).

446. C. S. Weickert *et al.*, Reduced brain-derived neurotrophic factor in prefrontal cortex of patients with schizophrenia. *Mol. Psychiatry*. **8**, 592–610 (2003).

447. K. Macdonald, D. Feifel, Oxytocin in schizophrenia: a review of evidence for its therapeutic effects. *Acta Neuropsychiatr.* **24**, 130–146 (2012).

448. A. Meyer-Lindenberg, G. Domes, P. Kirsch, M. Heinrichs, Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. *Nat. Rev. Neurosci.* **12**, 524–38 (2011).

449. C. Montag *et al.*, Oxytocin and oxytocin receptor gene polymorphisms and risk for schizophrenia: a case-control study. *World J. Biol. Psychiatry*. **14**, 500–8 (2013).

450. L. H. Rubin *et al.*, Peripheral oxytocin is associated with reduced symptom severity in schizophrenia. *Schizophr. Res.* **124**, 13–21 (2010).

451. D. Feifel *et al.*, Adjunctive intranasal oxytocin reduces symptoms in schizophrenia patients. *Biol. Psychiatry*. **68**, 678–80 (2010).

452. J. B. Deijen, M. I. Van Driel, M. L. Drent, The Involvement of the GH/IGF-I Axis in Cognitive Functions of Adult Patients and Healthy Subjects. *Open Endocrinol. J.* **6**, 68–79 (2012).

453. T. Biederer *et al.*, SynCAM, a synaptic adhesion molecule that drives synapse assembly. *Science*. **297**, 1525–31 (2002).

454. E. Díaz, SynDIG1 regulation of excitatory synapse maturation. *J. Physiol.* **590**, 33–8 (2012).

455. A. J. Funk, R. E. McCullumsmith, V. Haroutunian, J. H. Meador-Woodruff, Abnormal activity of the MAPK- and cAMP-associated signaling pathways in frontal cortical areas in postmortem brain in schizophrenia. *Neuropsychopharmacology*. **37**, 896–905 (2012).

456. D. Juraeva *et al.*, Integrated pathway-based approach identifies association between genomic regions at CTCF and CACNB2 and schizophrenia. *PLoS Genet.* **10**, e1004345 (2014).

457. M. Faizi, A. Salimi, M. Rasoulzadeh, P. Naserzadeh, J. Pourahmad, Schizophrenia induces oxidative stress and cytochrome C release in isolated rat brain mitochondria: a possible pathway for induction of apoptosis and neurodegeneration. *Iran. J. Pharm. Res. IJPR*. **13**, 93–100 (2014).

458. L. A. Glantz, J. H. Gilmore, J. A. Lieberman, L. F. Jarskog, Apoptotic mechanisms and the synaptic pathology of schizophrenia. *Schizophr. Res.* **81**, 47–63 (2006).

459. M. W. Rice, K. L. Smith, R. C. Roberts, E. Perez-Costas, M. Melendez-Ferro, Assessment of cytochrome C oxidase dysfunction in the substantia nigra/ventral tegmental area in schizophrenia. *PLoS One*. **9**, e100054 (2014).

460. L. Cavelier *et al.*, Decreased cytochrome-c oxidase activity and lack of age-related accumulation of mitochondrial DNA deletions in the brains of schizophrenics. *Genomics*. **29**, 217–24 (1995).

461. W. R. Perlman *et al.*, Alteration in estrogen receptor alpha mRNA levels in frontal cortex and hippocampus of patients with major mental illness. *Biol. Psychiatry*. **58**, 812–24 (2005).

462. P. Jia, L. Wang, H. Y. Meltzer, Z. Zhao, Common variants conferring risk of schizophrenia: a pathway analysis of GWAS data. *Schizophr. Res.* **122**, 38–42 (2010).

463. Y. H. Lee, J.-H. Kim, G. G. Song, Pathway analysis of a genome-wide association study in schizophrenia. *Gene*. **525**, 107–15 (2013).

464. C. S. Weickert *et al.*, Variants in the estrogen receptor alpha gene and its mRNA contribute to risk for schizophrenia. *Hum. Mol. Genet.* **17**, 2293–309 (2008).

465. S. H. Fatemi *et al.*, Chronic olanzapine treatment causes differential expression of genes in frontal cortex of rats as revealed by DNA microarray technique. *Neuropsychopharmacology*. **31**, 1888–99 (2006).

466. H. C. Galfalvy *et al.*, Sex genes for genomic analysis in human brain: internal controls for comparison of probe level data extraction. *BMC Bioinformatics*. **4**, 37 (2003).

467. B. Ji, K. K. Higa, J. R. Kelsoe, X. Zhou, Over-expression of XIST, the Master Gene for X Chromosome Inactivation, in Females With Major Affective Disorders. *EBioMedicine*. **2**, 909–918 (2015).

468. G. P. Otto, B. J. Nichols, The roles of flotillin microdomains--endocytosis and beyond. *J. Cell Sci.* **124**, 3933–40 (2011).

469. A. F. M. Moorman *et al.*, Patterns of Expression of Sarcoplasmic Reticulum Ca2+-ATPase and Phospholamban mRNAs During Rat Heart Development. *Circ. Res.* **76**, 616–625 (1995).

470. E. Tahara *et al.*, G1P3, an interferon inducible gene 6-16, is expressed in gastric cancers and inhibits mitochondrial-mediated apoptosis in gastric cancer cell line TMK-1 cell. *Cancer Immunol. Immunother.* **54**, 729–40 (2005).

471. A. Asai *et al.*, High Level Calcineurin Activity Predisposes Neuronal Cells to Apoptosis. *J. Biol. Chem.* **274**, 34450–34458 (1999).

472. P. G. Hogan, L. Chen, J. Nardone, A. Rao, Transcriptional regulation by calcium, calcineurin, and NFAT. *Genes Dev.* **17**, 2205–32 (2003).

473. H. Zeng *et al.*, Forebrain-specific calcineurin knockout selectively impairs bidirectional synaptic plasticity and working/episodic-like memory. *Cell*. **107**, 617–29 (2001).

474. T. Miyakawa *et al.*, Conditional calcineurin knockout mice exhibit multiple abnormal behaviors related to schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8987–92 (2003).

475. Z. Li *et al.*, Similarities in the behavior and molecular deficits in the frontal cortex between the neurotensin receptor subtype 1 knockout mice and chronic phencyclidine-treated mice: relevance to schizophrenia. *Neurobiol. Dis.* **40**, 467–77 (2010).

476. B. Kinkead, Neurotensin: an endogenous antipsychotic? *Curr. Opin. Pharmacol.* **2**, 99–103 (2002).

477. B. Kinkead, S. M. Lorch, M. J. Owens, C. B. Nemeroff, Sex- and Estrous Cycle-Related Differences in the Effects of Acute Antipsychotic Drug Administration on Neurotensin-Containing Neurons in the Rat Brain. *J. Pharmacol. Exp. Ther.* **295**, 205–211 (2000).

478. S. A. Boorjian *et al.*, Expression and significance of androgen receptor coactivators in urothelial carcinoma of the bladder. *Endocr. Relat. Cancer*. **16**, 123–37 (2009).

479. G. Pandini *et al.*, Androgens up-regulate the insulin-like growth factor-I receptor in prostate cancer cells. *Cancer Res.* **65**, 1849–57 (2005).

480. S. Hu, P. K. Peterson, C. C. Chao, Cytokine-mediated neuronal apoptosis. *Neurochem. Int.* **30**, 427–31.

481. S. L. Bronson, T. L. Bale, Prenatal stress-induced increases in placental inflammation and offspring hyperactivity are male-specific and ameliorated by maternal antiinflammatory treatment. *Endocrinology*. **155**, 2635–46 (2014).

482. T. Schneider *et al.*, Gender-specific behavioral and immunological alterations in an animal model of autism induced by prenatal exposure to valproic acid. *Psychoneuroendocrinology*. **33**, 728–40 (2008).

483. S. P. Hudson, S. Jacobson-Pick, H. Anisman, Sex differences in behavior and pro-inflammatory cytokine mRNA expression following stressor exposure and re-exposure. *Neuroscience*. **277**, 239–49 (2014).

484. O. Köhler *et al.*, Effect of Anti-inflammatory Treatment on Depression, Depressive Symptoms, and Adverse Effects. *JAMA Psychiatry*. **71**, 1381–91 (2014).

485. S. M. Sirmans, K. A. Pate, Epidemiology, diagnosis, and management of polycystic ovary syndrome. *Clin. Epidemiol.* **6**, 1–13 (2013).

486. E. Ingudomnukul, S. Baron-Cohen, S. Wheelwright, R. Knickmeyer, Elevated rates of testosterone-related disorders in women with autism spectrum conditions. *Horm. Behav.* **51**, 597–604 (2007).

487. S. Baron-Cohen, The extreme male brain theory of autism. *Trends Cogn. Sci.* **6**, 248–254 (2002).

488. N. Xita, A. Tsatsoulis, Review: fetal programming of polycystic ovary syndrome by androgen excess: evidence from experimental, clinical, and genetic association studies. *J. Clin. Endocrinol. Metab.* **91**, 1660–6 (2006).

489. D. A. Crain *et al.*, Female reproductive disorders: the roles of endocrine-disrupting compounds and developmental timing. *Fertil. Steril.* **90**, 911–40 (2008).

490. T. F. van de Mortel, Faking It: Social Desirability Response Bias in Self-report Research. *Aust. J. Adv. Nurs.* **25**, 40 (2008).

491. I. N. Karatsoreos, Links between Circadian Rhythms and Psychiatric Disease. *Front. Behav. Neurosci.* **8**, 162 (2014).

492. A. Reiner, D. Yekutieli, Y. Benjamini, Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*. **19**, 368–375 (2003).

493. Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).

494. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9440–5 (2003).