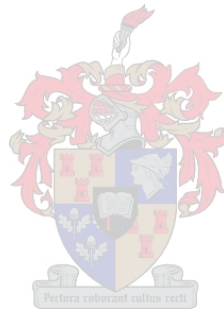


Investigating the functional significance of genome-wide variants associated with antipsychotic treatment response in schizophrenia

by
Ellen Susan Ovenden

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Genetics (Faculty of AgriSciences) at Stellenbosch University



Supervisor: Prof Louise Warnich
Co-supervisor: Prof Robin Emsley

December 2015

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights, and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2015

ABSTRACT

Schizophrenia is a debilitating disease affecting approximately 70 million people worldwide. Response to treatment, much like the disorder itself, is highly heritable, heterogeneous, and poorly understood. Only 50% of patients respond well to medication, and extensive research has provided limited improvement on this figure. Advances in genetic technologies coupled with massive increases in study sample size have the potential to explain the “missing heritability” of both schizophrenia and treatment response. Genome-wide association studies (GWAS) are at the forefront of complex trait research, but have had minimal success in terms of explaining the biology of psychiatric drug response. Despite the majority of GWAS “hits” being located in noncoding regions, functional interpretation is usually restricted to the closest gene. The Encyclopedia of DNA Elements (ENCODE) project has recently shown that noncoding variation is not just a functional proxy of adjacent coding regions, but can have complex and pervasive regulatory effects.

This study aimed to investigate the functionality of noncoding single nucleotide polymorphisms (SNPs) in schizophrenia treatment response. A novel bioinformatics pipeline incorporated coding and noncoding variants implicated in treatment response, regions of linkage disequilibrium (LD), regulatory data, and biological pathway predictions. Firstly, the literature was mined to identify all variants associated via GWAS with antipsychotic response, after which publically available data was employed to find markers in LD with these variants. This larger group of variants was analysed with bioinformatic tools such as RegulomeDB and rSNPBase to determine regulatory potential. Thereafter, affected gene targets and pathways were identified with DAVID and GeneMANIA. In order to investigate the findings further, the top predicted regulatory variants and their GWAS partners were genotyped with TaqMan® OpenArray® in a South African first episode schizophrenia (FES) cohort and analysed for associations with treatment outcomes.

The bioinformatic portion of this study implicated a region on chromosome 4q24 associated with treatment-refractory schizophrenia through involvement of the nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (*NFKB1*) gene. This gene is a master regulator involved in immunity and has over 200 gene targets. *NFKB1* and immune dysregulation have both previously been implicated in schizophrenia, pointing to a genetic overlap between schizophrenia risk and antipsychotic treatment response. The most significant variants in the association analyses occurred at the 4q24 locus, with rs230493 and rs3774959 significantly associated with poor response in the negative symptom domain ($P < 0.0001$). These findings suggest a genetic link between persistent negative symptoms and

treatment nonresponse. Additionally, a 14-variant haplotype containing these two polymorphisms was associated with 4.41% higher positive symptom severity.

Not only do these results validate the importance of the 4q24 region in antipsychotic response, but they emphasise the overlap of schizophrenia risk and drug response, and the potential role of genomic dysregulation in undesirable treatment outcomes. *NFKB1* and other associated genes should be studied in population-specific, replicative cohorts, in order to validate potential biomarkers of treatment response. This study illustrated the importance of thorough GWAS interpretation and inclusion of coding and noncoding variants to form biological hypotheses and better understand antipsychotic response.

OPSOMMING

Skisofrenie is 'n aftakelende siekte wat sowat 70 miljoen mense wêreldwyd raak. Behandlingsreaksie is, baie soos die siekte self, hoogs oorerflik en heterogeen, en word nog swak verstaan. Slegs 50% van pasiënte reageer goed op medikasie, en uitvoerige navorsing het slegs beperkte verbetering op hierdie syfer tot gevolg gehad. Vooruitgang in genetiese tegnologieë tesame met 'n geweldige toename in studie-steekproefgrootte kan potensieel die “ontbrekende erflikheid” van sowel skisofrenie as behandelingsreaksie verklaar. Genoom-wye assosiasiestudies (GWAS) is aan die voorpunt van komplekse kenmerknavorsing, maar het tot dusver minimale sukses ten opsigte van die verklaring van die biologie van psigiatriese middelreaksie gehad. Ondanks die feit dat die meerderheid GWAS-trefpunte in niekoderende streke voorkom, is funksionele interpretasie gewoonlik tot die naaste geen beperk. Die Ensiklopedie van DNS-elemente- (ENCODE-)projek het onlangs bewys dat niekoderende variasie nie net 'n funksionele sekondus van naasliggende koderende streke is nie, maar komplekse en deurdringende regulerende gevolge kan hê.

Hierdie studie was daarop gemik om die funksionaliteit van niekoderende enkel-nukleotied-polimorfismes (ENPs) in skisofreniebehandelingsreaksie te ondersoek. 'n Nuwe bioïnfomatika-pyplyn het koderende en niekoderende variante wat by behandelingsreaksie betrek word, streke van koppelingsdisekwilibrium (KD), reguleringsdata, en biologiese padvoorspellings geïnkorporeer. Eerstens is die literatuur ondersoek om alle variante te identifiseer wat via GWAS met antipsigotika-reaksie geassosieer word, waarna algemeen beskikbare data gebruik is om merkers in KD met hierdie variante te vind. Hierdie groter groep variante is met bioïnfomatika-hulpmiddels soos RegulomeDB en rSNPBase ontleed om reguleringspotensiaal te bepaal. Daarna is geïmpakteerde geenteikens en paaie met DAVID en GeneMANIA geïdentifiseer. Ten einde die bevindings verder te ondersoek, is die top-voorspelde reguleringsvariante en hul GWAS-vennote met TaqMan® OpenArray® in 'n Suid-Afrikaanse eerste-episode-skisofrenie-kohort gegenotipeer en vir assosiasies met behandelingsuitkomst ontleed.

Die bioïnfomatika-gedeelte van hierdie studie het 'n streek op chromosoom 4q24 geïmpliseer wat deur betrokkenheid van die geen nukleêre-faktor-kappa ligte polipeptied geen bevorderaar in B-selle 1 (*NFKB1*) met behandelingsweerstandige skisofrenie geassosieer word. Hierdie geen, 'n meester-reguleerder wat op immuniteit betrekking het, het meer as 200 geenteikens. *NFKB1* en immuundisregulering is albei vantevore by skisofrenie geïmpliseer, wat op 'n genetiese oorvleueling van skisofrenie-risiko en antipsigotika-behandelingsreaksie dui. Die mees beduidende variante in die assosiasie het

by die 4q24-lokus voorgekom, met rs230493 en rs3774959 wat albei beduidend met swak ná-behandelingsreaksie in die negatiewe-simptoom-domein geassosieer was ($P < 0.00001$). Hierdie bevindings dui op 'n genetiese verband tussen volhardende negatiewe simptome en niereaksie op behandeling. Daarbenewens is 'n 14-variant-haplotipe wat hierdie twee polimorfismes bevat met 'n 4.41% hoër graad positiewe simptome geassosieer.

Hierdie resultate staaf nie net die belangrikheid van die 4q24-streek in antipsigotika-reaksies nie, maar beklemtoon ook die oorvleueling van skisofrenie-risiko en middelreaksie, en die potensiële rol van genoom-disregulering in ongewenste behandelingsuitkomst. *NFKB1* en ander verwante gene moet in populasiespesifieke, repliseerbare kohorte bestudeer word ten einde potensiële biomerkers van behandelingsreaksie te staaf. Hierdie studie illustreer die waarde van deeglike GWAS-interpretasie en die insluiting van koderende en niekoderende variante om biologiese hipoteses te vorm en antipsigotika-reaksies beter te begryp.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the following people and institutions:

The National Research Foundation (NRF) and Stellenbosch University for financial support.

My supervisor, Prof Louise Warnich, for being a role model, and for her constant support, patience, and mentorship throughout my postgraduate studies.

My co-supervisor, Prof Robin Emsley, for his support and guidance regarding the clinical aspects of schizophrenia.

Dr Britt Drögemöller for being my mentor throughout my honours and masters studies.

Dr Nathaniel McGregor, whom I admire greatly, for his support and ability to answer any question I throw at him.

Ms Lundi Korkie for her conversation, guidance, and sense of humour.

The EONKCS team for patient recruitment, sample collection, and clinical data.

Mr Michael Klein and the University of Utah DNA Sequencing and Genomics Core Facility for genotyping.

Prof Lize van der Merwe for assistance with statistical analyses.

Dr Naveed Ishaque for his contribution to the development of the bioinformatics pipeline.

Prof Dana Niehaus for a sobering and educational glimpse of the lived experience of schizophrenia.

My mother and brother for supporting me unconditionally.

My housemates and friends for support and proofreading.

JK Rowling and Stephen Fry for the Harry Potter audiobooks, which got me through many long days and nights in the lab.

TABLE OF CONTENTS

LIST OF FIGURES.....	x
LIST OF TABLES.....	xi
LIST OF SYMBOLS AND ABBREVIATIONS	xii

CHAPTER 1: Introduction

1.1. The global burden of mental illness	1
1.2. Pharmacogenomics	2
1.3. Genetic diversity in South Africa	3

CHAPTER 2: Literature review

2.1. Schizophrenia.....	5
2.1.1. Symptoms and stages.....	5
2.1.2. Diagnosis.....	7
2.1.3. Risk factors.....	8
2.1.4. Genetics	9
2.2. Antipsychotic treatment of schizophrenia	12
2.2.1. Background	12
2.2.2. Adverse drug reactions	13
2.2.3. Treatment response	14
2.3. Antipsychotic pharmacogenomics.....	16
2.3.1. Background	16
2.3.2. Genome-wide association studies	17
2.4. Functional effects of genetic variation.....	19
2.4.1. Background	19
2.4.2. Noncoding variation	20
2.4.3. Recent bioinformatic developments.....	22
2.5. The South African context.....	24
2.6. Overview of the current study	26
2.6.1. Aim and objectives	26
2.6.2. Strategy	27

CHAPTER 3: Bioinformatic identification of potential regulatory variants associated with antipsychotic treatment response

3.1. Summary.....	29
3.2. Introduction.....	30
3.3. Materials and methods.....	31

3.3.1.	Data-mining	31
3.3.2.	Variants in linkage disequilibrium	32
3.3.3.	RegulomeDB analysis	33
3.3.4.	rSNPBase analysis	35
3.3.5.	Variants affecting binding motifs	35
3.3.6.	Nonsynonymous coding variants	36
3.3.7.	Affected genes and pathways	36
3.3.8.	Tissue-specific gene expression	37
3.4.	Results	37
3.4.1.	Antipsychotic response GWAS	37
3.4.2.	GWAS cohort ancestry and LD analyses	38
3.4.3.	RegulomeDB analysis	42
3.4.4.	rSNPBase analysis	43
3.4.5.	Variants affecting binding motifs	43
3.4.6.	Nonsynonymous coding SNPs	48
3.4.7.	Affected genes and pathways	48
3.4.8.	Tissue-specific gene expression	52
3.5.	Discussion	52
3.5.1.	Antipsychotic response GWAS	53
3.5.1.1.	<i>GWAS study design</i>	53
3.5.1.2.	<i>Significant GWAS findings</i>	56
3.5.2.	Predicted rSNPs and their genomic effects	56
3.5.2.1.	<i>Regions implicated in immunity</i>	56
3.5.2.2.	<i>Ubiquitous regulatory factors</i>	57
3.5.2.3.	<i>The 4q24 locus and NFKB1</i>	58
3.5.3.	Genes and pathways relevant to antipsychotic response	59
3.5.4.	Study limitations	60
3.6.	Conclusion	60

CHAPTER 4: Associations between predicted regulatory variants and antipsychotic treatment outcomes in a South African schizophrenia cohort

4.1.	Summary	63
4.2.	Introduction	64
4.3.	Materials and methods	65
4.3.1.	Patient samples	65
4.3.2.	SNP prioritisation	66
4.3.3.	SNP genotyping	67
4.3.4.	Statistical analyses	69

4.4.	Results	70
4.4.1.	Clinical outcomes	70
4.4.2.	SNP genotyping	70
4.4.3.	Haplotype analyses	72
4.5.	Discussion	74
4.5.1.	Clinical outcomes	74
4.5.2.	SNP genotyping and frequency comparisons	75
4.5.3.	Associations with treatment outcomes	75
4.5.3.1	<i>The 4q24 region</i>	76
4.5.3.2	<i>Refractoriness, remission, and early response</i>	78
4.5.3.3	<i>Metabolic outcomes</i>	79
4.5.4.	GWAS comparisons	80
4.5.5.	Study limitations	81
4.6.	Conclusion	82
CHAPTER 5: Conclusion and future perspectives		
5.1.	Conclusion	83
5.2.	Future perspectives	84
REFERENCES		86
ELECTRONIC SOURCES		107
APPENDIX A: Bioinformatics supplementary data		110
APPENDIX B: Association analyses supplementary data		118
APPENDIX C: Conference outputs		124

LIST OF FIGURES

Figure 2.1: Representation of the stages observed during the course of schizophrenia (Tandon <i>et al.</i> , 2009).....	6
Figure 2.2: The spectrum of psychiatric disorders, illustrating overlap between symptoms (Adam, 2013).....	8
Figure 2.3: Bar graph demonstrating the exponential growth of sample size and discovery of schizophrenia risk loci in genetic studies (Flint and Munafò, 2014).	10
Figure 2.4: Burdens and interventions as determinants of schizophrenia outcome (Tandon <i>et al.</i> , 2010).....	13
Figure 2.5: 18 studies demonstrating balanced outcomes of good vs. poor response to antipsychotic treatment (van Os and Kapur, 2009).....	16
Figure 2.6: The NHGRI GWAS Catalog: 17 traits, including drug response, with significantly associated SNPs ($P \leq 5 \times 10^{-8}$) across the genome, as of December 2013 (www.genome.gov/gwastudies/)..	18
Figure 2.7: Illustration of the various types of functional elements within the genome defined by ENCODE (Ecker <i>et al.</i> , 2012).....	23
Figure 3.1: Magnified view of the 4q24 genomic region on the UCSC Genome Browser with ENCODE data tracks (http://genome.ucsc.edu/ENCODE/).....	46
Figure 3.2: GeneMANIA network for affected genes in the HIV-1 Nef pathway according to DAVID (<i>ACTG1</i> , <i>NFKB1</i> and <i>RB1</i>), with related genes in grey.	50
Figure 3.3: GeneMANIA networks indicating genes in the chronic myeloid leukaemia pathway, and human CMV and MAPK pathways according to DAVID, with related genes in grey.....	51
Figure 4.1: Ancestry contributions from five populations in the SAC FES individuals (Drögemöller, 2013).	70
Figure 4.3: Frequency comparisons between the FES cohort and HapMap and 1000 Genomes populations.	71
Figure 4.2: Allelic discrimination plot for rs6427540. VIC® and FAM® relative dye intensities indicate genotype.....	71
Figure 4.4: Two haplotype blocks on chromosome four, designated by Haploview version 4.2 ($r^2 \geq 0.8$; $LOD \geq 3$) (Barrett <i>et al.</i> , 2005).....	72
Figure S1: Two haplotype blocks on chromosome four, designated by Haploview version 4.2 ($D' > 0.7 - > 0.98$) (Barrett <i>et al.</i> , 2005)..	119

LIST OF TABLES

Table 2.1: Schizophrenia symptom items measured by the PANSS (Kay <i>et al.</i> , 1987).	5
Table 2.2: Selected rSNPs associated with changes in expression of pharmacogenes.	21
Table 3.1: Population groups on SNAP (http://www.broadinstitute.org/mpg/snap/).	33
Table 3.2: RegulomeDB scoring system, with category 1 being most significant and category 6 least significant (Boyle <i>et al.</i> , 2012).	34
Table 3.3: Significant SNPs from antipsychotic pharmacogenomic GWAS identified by HuGE Navigator and the NHGRI GWAS Catalog.	39
Table 3.4: Ancestry breakdown of the four cohorts studied by relevant GWAS and corresponding SNP populations included for LD analysis (http://www.broadinstitute.org/mpg/snap/).	42
Table 3.5: Top predicted rSNPs from RegulomeDB with associated regulatory targets and effects (http://regulome.stanford.edu/).	44
Table 3.6: rSNPBase annotations for significant RegulomeDB SNPs, arranged by genomic position (http://rsnp.psych.ac.cn/).	47
Table 3.7: Top 10 SNPs predicted by TRAP to increase or decrease motif binding affinity significantly.	48
Table 3.8: Pathways identified by DAVID for affected genes (http://david.abcc.ncifcrf.gov/home.jsp).	49
Table 3.9: Brain- and liver-specific expression of ten most affected genes according to FANTOM5 (http://fantom.gsc.riken.jp/5/sstar).	52
Table 4.1: SNPs genotyped in the South African FES cohort, including predicted rSNPs and corresponding GWAS SNPs.	68
Table 4.2: Top significant SNP and haplotype associations with treatment outcomes in the FES cohort, with effect models and sizes indicated.	73
Table S1: Unique tRap motifs and number of associated SNPs for antipsychotic response and three control traits.	114
Table S2: 118 affected genes uploaded to DAVID and GeneMANIA for further analyses.	115
Table S3: SNPs genotyped by TaqMan® OpenArray® in the FES cohort.	118
Tables S4a – e: Haplotypes in the FES cohort from Haploview with frequencies ≥ 0.01	120
Table S5: Significant associations ($P < 0.05$) with treatment outcomes in the FES cohort. Highlighted associations survived Bonferroni correction.	121

LIST OF SYMBOLS AND ABBREVIATIONS

3'	3-prime end
α	Alpha
&	And
β	Beta
χ^2	Chi-square
©	Copyright
\$	Dollar
=	Equal to
>	Greater than/ nucleic acid substitution
\geq	Greater than or equal to
<	Less than
\leq	Less than or equal to
μ l	Microliters
%	Percentage
®	Registered trademark
\pm	Standard deviation
™	Trademark
A	Adenine/ Alanine
AA	African American
<i>ACTG1</i>	Actin, gamma 1 gene
<i>ADH7</i>	Alcohol dehydrogenase class 4 mu/sigma chain gene
ADRs	Adverse drug reactions
AIDS	Acquired immunodeficiency syndrome
AIMs	Ancestry informative markers
AIMS	Abnormal Involuntary Movement Scale
AIWG	Antipsychotic-induced weight gain
ASW	African ancestry in south western America
BBID	Biological Biochemical Image Database
BED	Browser Extensible Data
BMI	Body mass index
BPRS	Brief Psychiatric Rating Scale
BRCA1	Breast cancer type I susceptibility protein
C	Cytosine
c.	Mutation in coding DNA
CAGE	Cap analysis of gene expression
<i>CLMN</i>	Calmin (calponin-like, transmembrane) gene
CATIE	Clinical Antipsychotic Trials of Intervention Effectiveness
cDNA	Complementary DNA
CDCV	Common disease – common variant
CDRV	Common disease – rare variant
CEPH	Centre d'Etude du Polymorphisme Humain
CEU	Utah residents with European ancestry from the CEPH collection
CGI-I	Clinical global impression improvement scale
CGI-S	Clinical global impression severity scale
CHB	Han Chinese in Beijing, China
CHBJPT	Han Chinese in Beijing, China, and Japanese in Tokyo, Japan
CHD	Chinese in metropolitan Denver, Colorado
ChIP-Seq	Chromatin immunoprecipitation and sequencing
Chr	Chromosome

CI	Confidence interval
CMV	Cytomegalovirus
<i>CNTNAP5</i>	Contactin associated protein-like 5 gene
CNVs	Copy number variants
CUTLASS	Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia
CYP	Cytochrome P450 enzyme
<i>CYP</i>	Cytochrome P450 gene
<i>CYP1A2</i>	Cytochrome P450, family 1, subfamily A, polypeptide 2 gene
<i>CYP2C19</i>	Cytochrome P450, family 2, subfamily C, polypeptide 19 gene
<i>CYP2D6</i>	Cytochrome P450, family 2, subfamily D, polypeptide 6 gene
<i>CYP3A4</i>	Cytochrome P450, family 3, subfamily A, polypeptide 4 gene
<i>CYP3A5</i>	Cytochrome P450, family 3, subfamily A, polypeptide 5 gene
D	Aspartate
D'	normalised measure of allelic association (linkage disequilibrium measurement)
D ₂	Dopamine type 2
DALYs	Disability-adjusted life years
DAVID	Database for Annotation, Visualization and Integrated Discovery
DHS	DNase I hypersensitive site
DMEs	Drug-metabolising enzymes
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
<i>DRD2</i>	Dopamine receptor D2 gene
<i>DRD3</i>	Dopamine receptor D3 gene
<i>DRD4</i>	Dopamine receptor D4 gene
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders version four
DSM-5	Diagnostic and Statistical Manual of Mental Disorders version five
DUP	Duration of untreated psychosis
E	Glutamate
EA	European American
ENCODE	Encyclopedia of DNA Elements
EPS	Extrapyramidal side effects
eQTLs	Expression quantitative trait loci
<i>et al.</i>	<i>Et alii</i>
etc.	<i>Et cetera</i>
FANTOM5	Functional annotation of the mammalian genome 5
FDA	Food and Drug Administration
FDR	False discovery rate
FES	First episode schizophrenia
FGAs	First generation antipsychotics
G	Guanine
g.	Mutation in genomic DNA
GATA2	GATA binding protein 2
gDNA	Genomic DNA
GIH	Gujarati Indians in Houston, Texas
GRCh37	Genome Reference Consortium human genome build 37
GWAS	Genome-wide association study
H	Histidine
H3K27Ac	Acetylation of lysine 27 on histone H3
HDL	High-density lipoprotein

Hg19	Human genome version 19
HGMD	Human Gene Mutation Database
HGNC	HUGO Gene Nomenclature Committee
HIV	Human immunodeficiency virus
HREC	Human Research and Ethics Committee
HuGE	Human genome epidemiology
HUGO	Human Genome Organisation
HWE	Hardy-Weinberg equilibrium
ICD-10	International Classification of Diseases version 10
ID	Identification/ identifier
i.e.	<i>Id est</i>
Inc.	Incorporated
iPSC	Induced pluripotent stem cell
JPT	Japanese in Tokyo, Japan
KEGG	Kyoto Encyclopedia of Genes and Genomes
L	Leucine
LAI	Long-acting injectables
LD	Linkage disequilibrium
LMIC	Low- to middle-income countries
LOD	Logarithm of the odds (to the base 10)
LWK	Luhya in Webuye, Kenya
MAF	Minor allele frequency
MANBA	Mannosidase, beta A, lysosomal gene
MAPK	Mitogen-activated protein kinase
MEX	Mexican ancestry in California, Los Angeles
MHC	Major histocompatibility complex
MKK	Maasai in Kinyawa, Kenya
mmol/L	Millimoles per litre
n	Number of samples
N/A	Not applicable
Nef	Negative factor
<i>NFKB1</i>	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 gene
ng	Nanograms
NHGRI	National Human Genome Research Institute
NIMH	National Institute of Mental Health
<i>NRGN</i>	Neurogranin gene
NTC	No template control
OMIM	Online Mendelian Inheritance in Man
OPTiMiSE	Optimization of Treatment and Management of Schizophrenia in Europe
<i>P</i>	Probability
p.	page number
<i>PDE4D</i>	cAMP-specific phosphodiesterase 4D gene
PP-2	Polymorphism Phenotyping version 2
PANSS	Positive and Negative Syndrome Scale
PANTHER	Protein Analysis Through Evolutionary Relationships

PCR	Polymerase chain reaction
PDGF	Platelet-derived growth factor
PGC	Psychiatric GWAS Consortium
PGI	Patient global impression
PharmGKB	Pharmacogenomics Knowledge Base
PolyPhen-2	Polymorphism Phenotyping version 2
PWM	Position weight matrix
QTc	Corrected interval between Q and T wave in electrocardiogram
<i>r</i>	Log ratio value (TRAP measurement)
<i>r</i> ²	Squared correlation coefficient (linkage disequilibrium measurement)
<i>RB1</i>	Retinoblastoma 1 gene
RDoC	Research Domain Criteria
RNA	Ribonucleic acid
rSNPs	Regulatory SNPs
S	Serine
SAC	South African Coloured population
SANS	Scale for the Assessment of Negative Symptoms
SAPS	Scale for the Assessment of Positive Symptoms
SAS	Simpson-Angus Scale
SGAs	Second generation antipsychotics
<i>SHC1</i>	Src homology 2 domain containing (SHC) transforming protein 1 gene
SIFT	Sorting Intolerant from Tolerant
<i>SLAMF1</i>	Signalling lymphocyte activation molecule family member 1 gene
<i>SLCO1B1</i>	Solute carrier organic anion transporter family member 1B1 gene
SNAP	SNP Annotation and Proxy search
SNPs	Single nucleotide polymorphisms
SSTAR	Semantic catalogue of Samples, Transcription initiation And Regulators
sTRAP	TRAP tool for analysis of single nucleotide changes
T	Thymine
<i>TCF4</i>	Transcription factor 4 gene
TD	Tardive dyskinesia
TF	Transcription factor
TFBS	Transcription factor binding site
<i>TJP1</i>	Tight junction protein 1 gene
TNF	Tumour necrosis factor
<i>TNFRFS11A</i>	TNF receptor superfamily, member 11a, NFKB activator gene
TPM	Tags per million
TRAP	Transcription factor Affinity Prediction
tRap	R package of TRAP
TSI	Toscans in Italy
UCSC	University of California, Santa Cruz
UDP	Uridine diphosphate
<i>UGT1A1</i>	UDP glucuronosyltransferase 1 family, polypeptide A1 gene
USA	United States of America
UTR	Untranslated region
<i>VKORC1</i>	Vitamin K epoxide reductase complex, subunit 1 gene
vs.	Versus

YLDs
YRI

Years lived with disability
Yoruba in Ibadan, Nigeria

ZNF202
ZNF804A

Zinc finger protein 202 gene
Zinc finger protein 804A gene

CHAPTER 1: Introduction

1.1. The global burden of mental illness

Psychiatric disorders place an immense burden on individuals, families, and communities. Worldwide, the combination of high prevalence, high cost of treatment, and high disability has long called for mental health to be prioritised in public health care (Murray and Lopez, 1996). Psychiatric illness constitutes approximately 13% of the global disease burden (World Health Organization, 2008). In 2010, mental and substance use disorders caused the fifth highest number of disability-adjusted life years (DALYs), according to the latest Global Burden of Disease study. In fact, these disorders were the leading cause of years lived with disability (YLDs), a subcategory of the DALY (Whiteford *et al.*, 2013). Compounding the problem, psychiatric disorders often demonstrate comorbidity with other chronic medical conditions, and can significantly worsen a patient's outcome (Patel *et al.*, 2013).

These findings have far-reaching consequences. Firstly, mental illness creates a global economic burden currently estimated at \$2.5 trillion, which is predicted to increase almost three-fold by 2030 (Bloom *et al.*, 2011). A major contributor to these costs is a lack of successful preventions and cures, resulting in relapse and hospitalisation (Ascher-Svanum *et al.*, 2010; Collins *et al.*, 2011). Secondly, stigmatisation of psychiatric disorders produces a large socioeconomic burden in both urban and rural settings. Affected individuals are often cut off from their community, thus restricted from health care, education, employment and social support, resulting in significantly shorter lifespans compared to the general population (Kadri and Sartorius, 2005).

Despite the serious and diverse problems created by mental illness, most countries do not allocate sufficient resources to psychiatric treatment (Saxena and Skeen, 2012). Mental well-being is not globally prioritised in comparison to other illness: according to the Mental Health Atlas, governments spend approximately \$2 per person on mental health annually (World Health Organization, 2011a). Even when an effective treatment strategy exists, it may not be implemented within the healthcare system, due to a lack of qualified staff or budgetary constraints (Tomlinson *et al.*, 2009). For the most part, individuals with mental disorders are treated in primary healthcare facilities, with only 20% of adults with common psychiatric problems in the United States of America (USA) consulting a mental health specialist (Wang *et al.*, 2005). Lack of proper care increases complications, stigma, and the already high costs associated with mental well-being (World Health Organization, 2011b).

Not surprisingly, the burden of mental illness is amplified in low- to middle-income countries (LMIC) such as South Africa. There is extensive evidence that poverty is strongly associated with increased risk for mental disorders (Patel and Kleinman, 2003; Murali and Oyeboode, 2004). Despite LMIC taking on as much as 75% of the global burden of psychiatric illness, their general healthcare budgets are lower, each with an even smaller portion dedicated to mental health (World Health Organization, 2011b). Demyttenaere and colleagues (2004) estimate that fewer than 24% of affected people in LMIC receive treatment. Furthermore, 70% of African countries allocate less than 1% of their healthcare budgets to mental health (Lund *et al.*, 2010).

The dire situation in LMIC is partly attributable to a lack of resources in the form of healthcare professionals. For example, there is an average of only one psychiatrist per two million individuals in low-income countries (Saxena and Skeen, 2012). To put this in perspective, the number of psychiatrists on the African continent is less than the number in the state of Massachusetts in the USA (Patel *et al.*, 2013). Additionally, mental health is deprioritised in LMIC due to high rates of other diseases such as HIV/AIDS and tuberculosis (Lund *et al.*, 2010). The immense health, socioeconomic, and financial burdens of psychiatric illness call for increased research, education, and healthcare resources, particularly in LMIC. Improving the understanding and treatment of these disorders is vital for ensuring sustainable mental well-being.

1.2. Pharmacogenomics

An important consideration for the treatment of any disease is pharmacogenomics, or the effect of genetic variation on drug response. Often, immense heterogeneity is seen in individuals treated with the same medication. This is largely influenced by variants in individuals' DNA, particularly in drug metaboliser and transporter genes (Ozomaro *et al.*, 2013; Carr *et al.*, 2014). In most cases, psychiatric drug treatment is standardised for all patients, proceeds by trial-and-error, and dose or medication type is adjusted only after a positive outcome is not reached (Cacabelos *et al.*, 2011). This is a costly and potentially dangerous exercise for the treatment of any disease, as drug toxicity and side effects are a reality for many patients. For example, Nyakutira and colleagues (2008) discovered that 50% of African patients receiving efavirenz for HIV treatment had blood concentrations above the toxicity threshold, as a result of a gene-dose interaction. With reference to psychiatric treatment, the administration of common antipsychotics can cause tardive dyskinesia, a chronic and severe movement disorder, in up to 30% of patients (Chowdhury *et al.*, 2011).

Psychiatric treatment is complex, chronic, and requires close monitoring of patients. Although expensive, standardised treatment is currently substantially cheaper than the resources required for personalised medicine. However, the implementation of pharmacogenomics in psychiatry is expected to reduce costs associated with long-term treatment outcomes. This field of research has the potential to minimise the development of side effects, treatment complications, and hospitalisations, ultimately lowering the amount of YLDs and DALYs associated with disease (León-Cachón *et al.*, 2012). In fact, pharmacogenomics has already demonstrated its ability to save money on disease treatment. Recently, pharmacogenetic screening of patients prior to treatment with the anti-cancer drug, trastuzumab, decreased the length of the clinical trial by approximately eight years, and saved millions of dollars (Cook *et al.*, 2009). Further demonstrating the importance of pharmacogenomic considerations, many drugs approved by the Food and Drug Administration (FDA) contain labels with pharmacogenomic indications, including over 30 psychiatric medications such as antipsychotics and antidepressants (<http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm>).

Despite these findings, pharmacogenomic applications are limited, as the majority of large-scale genetic studies have focused on disease susceptibility rather than treatment response. By investigating pharmacogenomic interactions in psychiatry, our understanding of treatment outcomes, and subsequently our ability to tailor treatment to the individual and improve drug design, will increase. The coupling of well-characterised clinical data with genetic and bioinformatic resources has great potential for alleviating the extensive burden placed on those with mental illness. This is particularly important in LMIC given the magnified burden of disease in these countries. Thus, pharmacogenomics is an essential starting point in the improved treatment of psychiatric diseases.

1.3. Genetic diversity in South Africa

Although pharmacogenomic research shows great promise, the overwhelming majority of studies is conducted in developed countries. Paradoxically, Hinds and colleagues (2005) estimate that LMIC contain up to 90% of human genetic variation, thus providing an unparalleled resource for genetic studies of complex disorders. In fact, Southern African populations have demonstrated the highest level of genetic diversity worldwide (Campbell and Tishkoff, 2008). The South African Coloured (SAC) population, for example, is highly admixed, with African, Asian, and European ancestry contributions (de Wit *et al.*, 2010; Daya *et al.*, 2013). South Africa therefore provides a rich genetic resource for uncovering the

architecture of complex traits (Ramsay, 2012), and should be viewed as an opportunity for genomic research rather than a disease burden to the world.

Nevertheless, South African individuals remain understudied and underrepresented in pharmacogenomic research (Drögemöller *et al.*, 2011). Indeed, the extreme gap between needs and available services in LMIC is mirrored by the so-called “10/90 gap” in research. This is the phenomenon that only 10% of global research funding is spent on the problems faced by the poorest 90% of the population (Global Forum for Health Research, 2000). Furthermore, only 5% of research published in high impact psychiatric journals originates from LMIC, with only 1% from South African authors (Patel and Sumathipala, 2001; de Jesus Mari *et al.*, 2009). There is no doubt that South Africa is home to unique and heterogeneous genetic variation, and clinically actionable findings from high-income countries may not be applicable. Therefore, increased study of its populations is vital for identifying the genetic differences underlying complex phenotypes such as psychiatric illness and treatment response. By combining the latest technological advances in genetics with overburdened and understudied ethnic groups, novel insights into psychiatric pharmacogenomics and improved treatment become possible.

CHAPTER 2: Literature review

2.1. Schizophrenia

2.1.1. Symptoms and stages

Schizophrenia is arguably the most debilitating psychiatric disorder, and consequently is highly stigmatised and costly to treat (van Os and Kapur, 2009). Indeed, of all the mental disorders investigated by the latest Global Burden of Disease study, schizophrenia accounted for the most disability (Whiteford *et al.*, 2013). The disorder is complex and pervasive, permeating all aspects of an individual's life and manifesting as a range of symptoms. Positive or psychotic symptoms are defined as exaggerated states of functioning, which are absent in the general population but present in schizophrenia, whilst negative symptoms constitute loss of a range of functions that are usually present in healthy individuals (Tandon *et al.*, 2009). For example, individuals with schizophrenia may experience hallucinations and delusions on the one hand, but impairments in speech, motivation and social interest, on the other. General psychopathological symptoms also occur, which include mood, motor and cognitive deficits. These symptoms can be quantified by different scales, the most common of which is the Positive and Negative Syndrome Scale (PANSS; Kay *et al.*, 1987). Seven items on this scale measure positive and negative symptoms, respectively, and 16 items measure general psychopathology, as shown in Table 2.1. Each of the 30 items on the test is scored from 1-7, increasing in severity. Therefore the baseline PANSS score is 30, and the maximum possible score is 210.

Table 2.1: Schizophrenia symptom items measured by the PANSS (Kay *et al.*, 1987).

Positive symptoms	Negative symptoms	General symptoms
Delusions	Blunted affect	Somatic concern
Conceptual disorganisation	Emotional withdrawal	Anxiety
Hallucinatory behaviour	Poor rapport	Guilt feelings
Grandiosity	Social withdrawal	Tension
Excitement	Difficulty in abstract thinking	Mannerisms and posturing
Suspiciousness	Lack of spontaneity	Depression
Hostility	Stereotyped thinking	Motor retardation
		Unusual thought content
		Uncooperativeness
		Disorientation
		Poor attention
		Lack of judgment and insight
		Poor impulse control
		Preoccupation
		Disturbance of volition
		Active social avoidance

The PANSS is widely used to determine symptom severity, response to treatment, relapse, and remission in schizophrenia (Levine *et al.*, 2011). Other scales include the Scales for the Assessment of Negative (SANS) and Positive (SAPS) Symptoms (Andreasen, 1983; 1984) and the Brief Psychiatric Rating Scale (BPRS) (Overall and Gorham, 1962).

Schizophrenia is a chronic disorder that typically displays a gradual deterioration in functioning. It can be divided into four stages or phases, indicated in Figure 2.1. Generally, negative and cognitive symptoms surface in childhood or adolescence, followed by the development of psychotic symptoms in young adulthood (Mueser and McGurk, 2004). The first psychotic episode marks the beginning of the psychotic phase and the official onset of schizophrenia, which is usually followed by subsequent episodes in between brief periods of remission (Lieberman *et al.*, 2001). The disorder then reaches a stable plateau, which is characterised by residual negative and cognitive symptoms and a general decline in functioning (Tandon *et al.*, 2009).

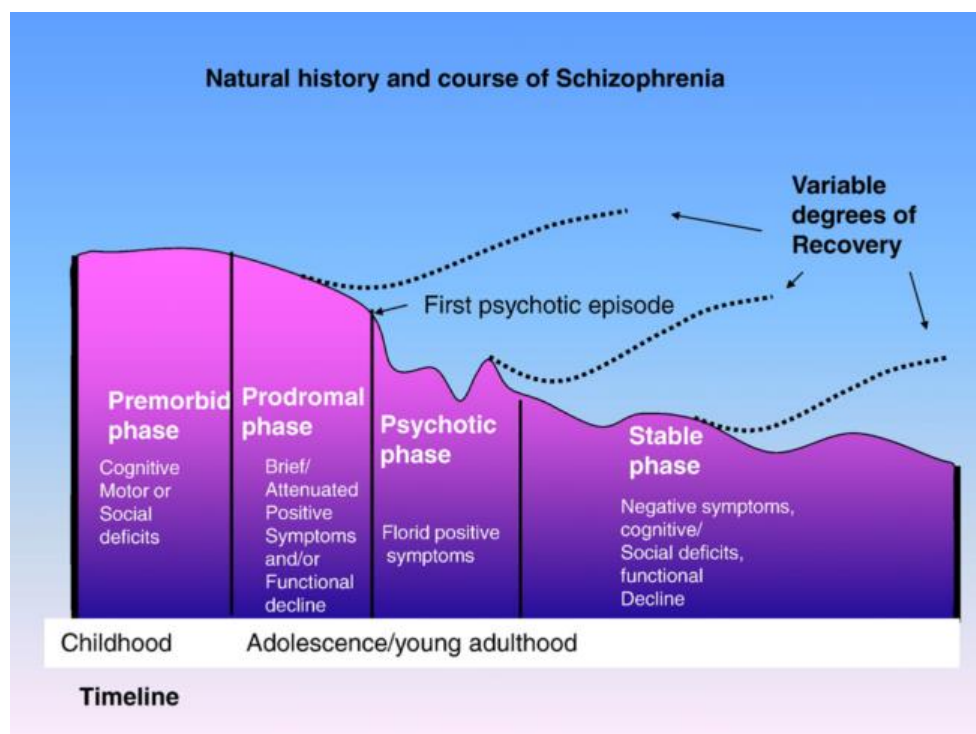


Figure 2.1: Representation of the stages observed during the course of schizophrenia (Tandon *et al.*, 2009). *Reprinted with permission from Elsevier.*

Despite the classification of schizophrenia into different stages, diagnosis of the disorder is difficult. Tandon and colleagues (2009) discuss several limitations of the four-phase model of schizophrenia. Firstly, there is extensive heterogeneity in the type and severity of symptoms seen in individuals, making differentiation between phases difficult. Psychotic symptoms

often do not manifest in clear intervals, therefore the definition of the first episode of psychosis is somewhat arbitrary. Additionally, more than half of patients that experience mild positive symptoms in the prodromal stage do not go on to develop the disorder. Lastly, the time course of the illness and extent of deterioration vary between patients (Tandon *et al.*, 2009). Nevertheless, relapses and persistence of symptoms despite treatment create a chronic struggle with schizophrenia for the majority of individuals (Albus, 2012).

2.1.2. Diagnosis

The current diagnosis of schizophrenia is determined by clinical interview, based on criteria either in the fifth version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013), or the International Classification of Diseases version 10 (ICD-10; World Health Organization, 2015), which are similar and display high diagnostic reliability (Peralta and Cuesta, 2003; Mueser and McGurk, 2004). The most commonly used system, the DSM-5, advises diagnosis when an individual exhibits two or more core symptoms, i.e. hallucinations, delusions, negative symptoms, or disorganised thinking. In addition, these symptoms must be present for at least a month before a patient can be diagnosed as experiencing their first psychotic episode (American Psychiatric Association, 2013). In contrast to previous versions, the DSM-5 does not divide schizophrenia into subtypes (paranoid, catatonic, disorganised, schizoaffective, undifferentiated, and residual), as this approach has shown limited reliability and validity, and poor clinical success (Tandon, 2014). Instead, the manual proposes a broad and thorough assessment of symptom severity to address the substantial variation that exists between patients.

The heterogeneity of schizophrenia poses another problem to diagnosis: there is extensive overlap with other psychiatric disorders. On the whole, research does not support the compartmentalisation of these disorders, as most mental illnesses have been found to share risk factors, symptoms, and biological pathways (Adam, 2013; Doherty and Owen, 2014). This is displayed in Figure 2.2, in which psychiatric disorders lie upon a spectrum. There is a need for reconsideration of nosological boundaries, as many researchers agree that schizophrenia's heterogeneity means it should not be defined as a single disease (Tandon, 2012; Barch *et al.*, 2013; Alvarez-Rodriguez *et al.*, 2014; Arnedo *et al.*, 2014). Although the DSM-5 does not address this developing paradigm shift, taking a dimensional approach to diagnosis is a promising first step towards an improved understanding of this complex disorder. To provide more precise diagnosing in psychiatry, the National Institute of Mental Health (NIMH) has developed the Research Domain Criteria (RDoC), which shifts focus away from symptoms onto biologically distinct psychopathological mechanisms (Insel *et al.*,

2010; Insel and Cuthbert, 2015). Classification of patient subgroups with RDoC considers specific biosignatures, identifiable through genetic research and neuroimaging (Insel *et al.*, 2010). Studies implementing this method are few and require validation, but this is a promising step in improving schizophrenia diagnosis and outcome.

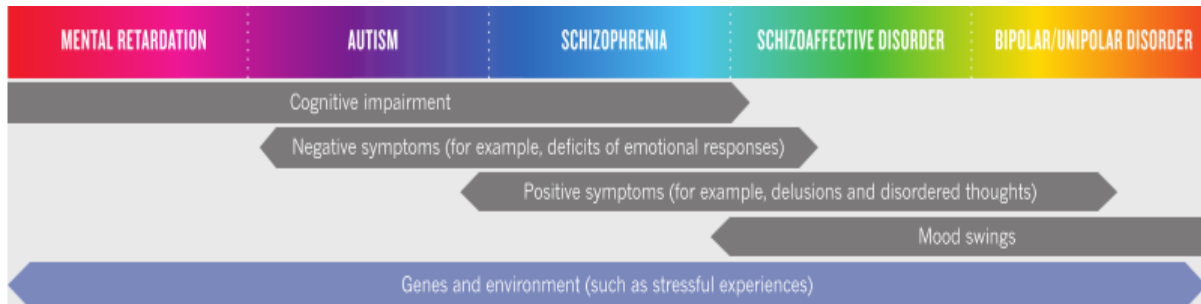


Figure 2.2: The spectrum of psychiatric disorders, illustrating overlap between symptoms (Adam, 2013). *Reprinted with permission from Nature Publishing Group.*

2.1.3. Risk factors

Schizophrenia presents a lifetime risk of 0.7% (Tandon *et al.*, 2008), with a prevalence of up to 1% in the general population (Curtis, 2013). Its aetiology and biological mechanisms are poorly understood, and much like the other features of schizophrenia, the risk factors for development of the disorder are heterogeneous. The establishment and severity of schizophrenia involve the interplay between several genetic and environmental influences (Tsuang *et al.*, 2004; Singh *et al.*, 2014).

Pre- and perinatal risk factors for schizophrenia include maternal infection, stress, malnutrition and obstetric complications (Opler *et al.*, 2013). Individuals that have experienced childhood trauma also show increased risk for the disorder (Schmitt *et al.*, 2014). Various sociodemographic stressors contribute towards schizophrenia, such as urbanicity (Krabbendam and van Os, 2005), migration (Cantor-Graae and Selten, 2005) and lower social class (Mueser and McGurk, 2004). The “social defeat hypothesis” suggests that occupying a lower social standing or belonging to a minority increases risk for the disorder (van Os *et al.*, 2010). Finally, cannabis use has been linked to schizophrenia development (van Os and Kapur, 2009). Despite environmental elements, it is widely acknowledged that genetic predisposition is the top contributing risk factor for schizophrenia, with family history of the disorder being the most reliable predictor for development in an individual (Sullivan, 2005; Clarke *et al.*, 2012).

2.1.4. Genetics

Family, adoption and twin studies have shown the heritability of schizophrenia to be approximately 81%, making it one of the most heritable psychiatric disorders (Sullivan *et al.*, 2003; Singh *et al.*, 2014). The risk of developing the disorder increases with the degree of relatedness to an affected individual. For example, the concordance between monozygotic twins is three times greater than between dizygotic twins (Clarke *et al.*, 2012, Girard *et al.*, 2012). Despite evidence of genetic aetiology, the complex, non-Mendelian nature of schizophrenia and other psychiatric disorders has made the exact biological underpinnings tricky to elucidate (Singh *et al.*, 2014).

Currently, there are two major hypotheses with regards to the genetic mechanisms of schizophrenia. The common disease – common variant (CDCV) hypothesis proposes that many commonly occurring genomic variants of small effect size bring about a cumulative increase in schizophrenia susceptibility. Conversely, the common disease – rare variant (CDRV) hypothesis states that a small number of rare, but highly penetrant variants of large effect size confer the majority of schizophrenia risk (Stefansson *et al.*, 2009; van Dongen and Boomsma, 2013). Recent findings suggest that the truth lies somewhere between these two, with a combination of heterogeneous rare and common alleles culminating in the pathophysiology of the disease (Mowry and Gratten, 2013).

Genetic research has unveiled extensive results across this spectrum of variants. Earlier studies relied on linkage analyses, which look at co-segregating variants in families, and can be a successful tool for understanding simple Mendelian diseases (Kerem *et al.*, 1989; Muir *et al.*, 1995; Mowry and Gratten, 2013). Poor replication and weak significance signals led to the abandonment of this approach, in favour of a more complex, polygenic view of schizophrenia (Rodriguez-Murillo *et al.*, 2012). Candidate gene association studies were the next advancement in the study of schizophrenia. This method compares a particular gene in schizophrenia cases and controls and determines whether there are common variants that associate with the disease (Kim *et al.*, 2011). The SzGene database is a record of all genetic association studies, and contains over a thousand genes studied with the candidate approach (Allen *et al.*, 2008). However, many results are inconsistent and the majority of studies have not been replicated. Additionally, this hypothesis-bound method is restrictive, since selecting a candidate gene is based on the limited knowledge we have of schizophrenia (Collins *et al.*, 2012).

Almost a decade ago, fuelled by advances in genotyping technology, the first genome-wide association study (GWAS) on schizophrenia was performed (Mah *et al.*, 2006). GWAS have

significant advantages over previous study designs. Firstly, they do not require selection of candidate genes; in other words they provide an unbiased and hypothesis-free approach, creating the potential for discovery of novel schizophrenia loci (Zhang and Malhotra, 2013a). Secondly, by scanning the entire genome, GWAS can simultaneously analyse millions of single nucleotide polymorphisms (SNPs) and determine association with schizophrenia in large case/ control groups (Kim *et al.*, 2011).

This approach gives enormous support to the CDCV hypothesis. Since GWAS have been applied to the field of schizophrenia, over 100 independent variants have been identified in more than 15 GWAS, in unprecedented sample sizes (Zhang and Malhotra, 2013a; McCarthy *et al.*, 2014). This highlights the importance of large sample sizes, with increased sample size leading to more associations (Figure 2.3). The most notable findings that have been replicated in subsequent studies are variants in the zinc finger protein 804A (*ZNF804A*) gene, the major histocompatibility complex (MHC) genes, the neurogranin (*NRGN*) gene, the transcription factor 4 (*TCF4*) gene, and the dopamine receptor D2 (*DRD2*) gene (Rodriguez-Murillo *et al.*, 2012; Ripke *et al.*, 2014). The MHC locus is currently the most replicated finding, suggesting a role for the immune system in schizophrenia development (Sullivan *et al.*, 2012; Ripke *et al.*, 2014).

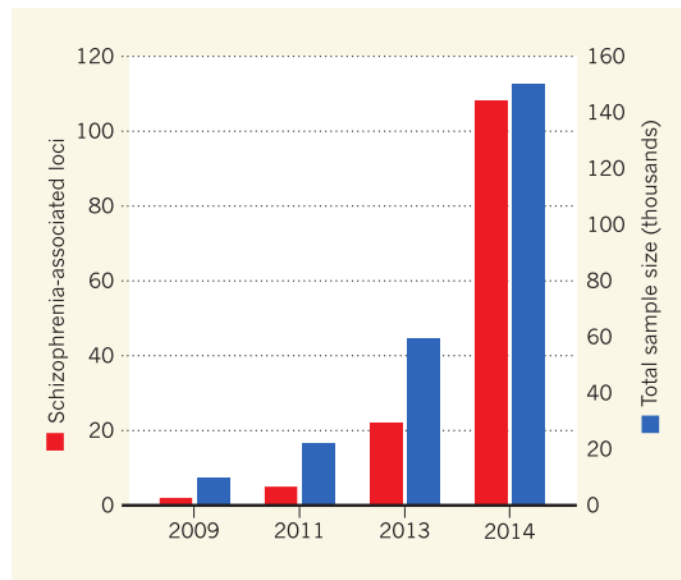


Figure 2.3: Bar graph demonstrating the exponential growth of sample size and discovery of schizophrenia risk loci in genetic studies (Flint and Munafò, 2014). *Reprinted with permission from Nature Publishing Group.*

The most noteworthy contributor to this field is the Psychiatric GWAS Consortium (PGC), which is spread across 19 countries and over 60 institutions, and currently has access to about 40 000 genomes for the study of schizophrenia (Sullivan, 2010; Wright, 2014). With

the aim of performing large-scale analyses of psychiatric disorders, the PGC has yielded a plethora of results relevant to schizophrenia. Their most recent study identified 108 significant risk loci, 83 of which were novel. However, there were significant results for genes involved in neurotransmitter systems, such as *DRD2*, which are consistent with previous hypotheses of impaired neurodevelopmental functioning in schizophrenia (Ripke *et al.*, 2014).

The growing number of novel loci for schizophrenia susceptibility suggests that the disorder is even more complex than previously assumed. Moreover, it is estimated that these common alleles only account for 1-2% of genetic risk for schizophrenia, making them neither vital nor sufficient for development of the disorder (Zhang and Malhotra, 2013a). One must also consider that GWAS have limitations. Firstly, the nature of multiple testing requires independent replication studies to ensure that variants are not simply statistical artefacts (Bertram, 2008), but the majority of GWAS “hits” have not been successfully replicated (Sham and Purcell, 2014). Secondly, there is a lack of post-GWAS functional analyses of significant loci, leading to a growing list of potentially important genomic regions, but minimal understanding of how they operate (Girard *et al.*, 2012; refer to 2.3.2. for more about GWAS in relation to the current study).

The case of missing heritability may in part be solved by analysing rare variants of large effect, as stipulated by the CDRV hypothesis. Copy number variants (CNVs) are rare mutations that are highly penetrant and demonstrate large effect sizes (Zhang and Malhotra, 2013a). The most notable example is a *de novo* microdeletion on chromosome 22q11.2, with carriers exhibiting a three-fold increase in risk for schizophrenia (Sullivan *et al.*, 2012). Rare point mutations have also been implicated in schizophrenia, although this type of study is in its infancy (Mowry and Gratten, 2013). With advances in whole-genome and whole-exome sequencing, Xu *et al.* (2011) have shown that protein-altering *de novo* mutations are enriched in individuals with schizophrenia, which was confirmed in an independent study by Girard and colleagues (2011). More recently, a large study sequenced exomes of over 5000 individuals, and found rare mutations across many genes that were significantly associated with schizophrenia (Purcell *et al.*, 2014).

These findings highlight the importance of rare variants in future studies of schizophrenia. A few years ago, the Grand Challenges in Global Mental Health initiative listed the identification of biomarkers as one of the top 25 challenges for progress in mental health (Collins, 2011). Ideally, genetic features of schizophrenia should be incorporated into its clinical conceptualisation and diagnosis, and there is a call for a more “biologically relevant”

nosology (Tandon, 2012; Kim and State, 2014). The current debate about the missing heritability of the disorder has generated progress in the form of many heterogeneous risk loci. It has been proposed that diverse, large-scale techniques in combination with functional analyses be used to identify the remaining predictors across the risk spectrum (Mowry and Gratten, 2013). This approach has the potential to improve our understanding of this complex disorder.

2.2. Antipsychotic treatment of schizophrenia

2.2.1. Background

The treatment of schizophrenia was revolutionised with the chance discovery of chlorpromazine's antipsychotic properties in the 1950s (Lopez-Munoz *et al.*, 2005). Carlsson and Lindqvist (1963) subsequently determined that this drug's success was brought about by dopamine receptor antagonism. This marked the establishment of the dopamine hypothesis in schizophrenia treatment (Kapur and Mamo, 2003). Today, over 60 years since the introduction of chlorpromazine, all antipsychotics include dopamine D₂ receptor blockade in their mechanism of action (Brandl *et al.*, 2014).

Chlorpromazine was the first of over 60 antipsychotics designed to treat schizophrenia (Tandon *et al.*, 2010). These drugs can be divided into two classes: the earlier, typical, or first generation antipsychotics (FGAs), and the more recent, atypical, or second generation antipsychotics (SGAs). Overall, studies have shown that FGAs effectively reduce psychotic symptoms and prevent relapses in schizophrenia, but other symptoms persist (Arranz and de Leon, 2007; Carpenter and Davis, 2012). These lingering negative and cognitive deficits contribute largely to general functional decline and long-term decreased quality of life (Kirkpatrick *et al.*, 2006). The introduction of SGAs sought to improve upon treatment outcomes by incorporating a wider range of neurochemical targets than FGAs. Besides the D₂ receptor, SGAs act on other components of the dopaminergic pathway, as well as the serotonergic, glutamatergic and adrenergic systems (Meltzer, 2013).

Despite their multi-target profile, there is ongoing debate about whether SGAs offer treatment advantages over FGAs. Only a handful of large-scale studies comparing effectiveness of antipsychotics have been performed, including the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE; Lieberman *et al.*, 2005) and the Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia (CUtLASS; Jones *et al.*, 2006). Both of these studies found no significant differences in the efficacy between the two generations of antipsychotics, but there were notable flaws in their study designs (Meltzer, 2013). The only

atypical antipsychotic that has clear, extensively replicated advantages over typical antipsychotics is clozapine, which is highly successful in treatment-refractory schizophrenia when other drugs fail (McEvoy *et al.*, 2006; Bonham and Abbott, 2008; Chowdhury *et al.*, 2011). There is still much to learn about the mechanisms of these drugs, and advances in drug design have been relatively modest (Carpenter and Davis, 2012). Other treatment options for schizophrenia are illustrated in Figure 2.4. Despite antipsychotics being the most effective option, combining them with other forms of treatment is necessary for improved quality of life, given the complex and often lifelong nature of the disorder (Tandon *et al.*, 2010).

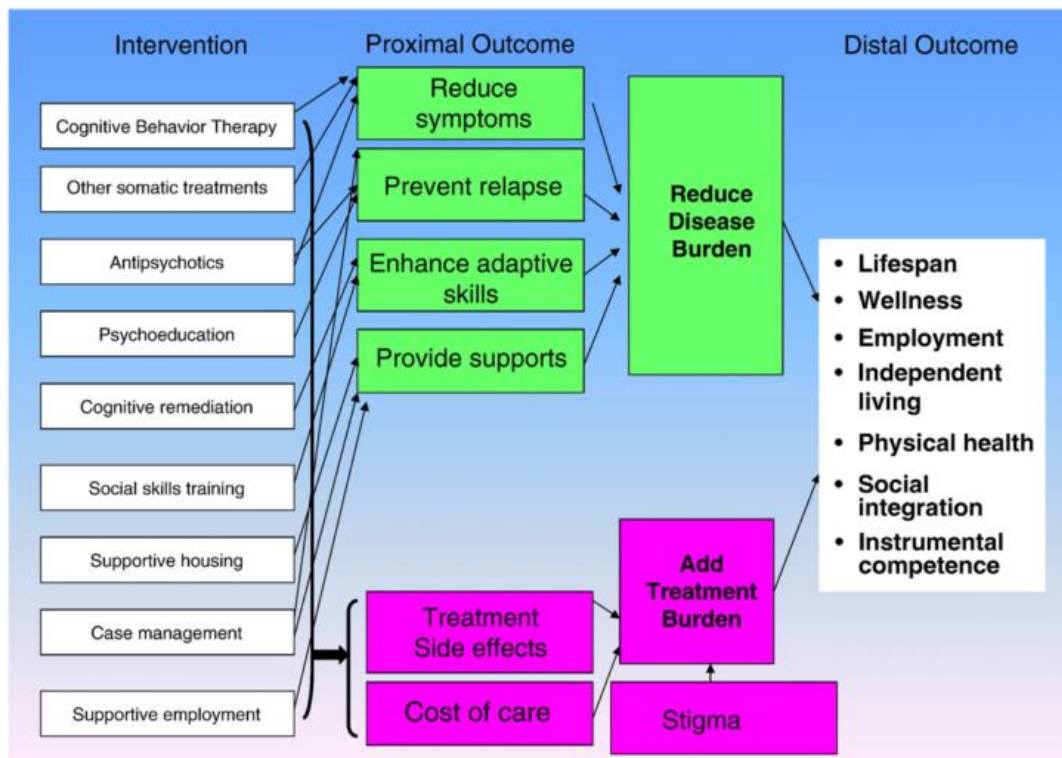


Figure 2.4: Burdens and interventions as determinants of schizophrenia outcome (Tandon *et al.*, 2010). Reprinted with permission from Elsevier.

2.2.2. Adverse drug reactions

Perhaps the most apparent distinction between FGAs and SGAs is the different adverse drug reactions (ADRs) with which they are associated. Generally, ADRs caused by antipsychotics are diverse, severe, and can be long-lasting (Zandi and Judy, 2010). FGAs are associated with motor abnormalities, such as acute and reversible extrapyramidal side effects (EPS), namely dystonia, akathisia, and parkinsonism, or with chronic conditions, such as tardive dyskinesia (TD; Tandon *et al.*, 2010). TD is the most extensively studied ADR and

occurs in 20-30% of individuals after three months of treatment with FGAs (Chowdhury *et al.*, 2011).

In contrast, SGAs present a significantly lower risk of EPS and are predominantly linked to weight gain and other metabolic side effects (Tandon *et al.*, 2010). Antipsychotic-induced weight gain (AIWG) is observed in up to 30% of SGA-treated patients. Additionally, selected SGAs increase the risk of cardiac complications, such as the prolongation of the QT interval (Brennan, 2014). The uniqueness of clozapine applies to its side effect profile as well as its effect on treating nonresponse in schizophrenia. It has been associated with a small but life-threatening risk of agranulocytosis, a condition characterised by a decrease in neutrophil count (Alvir *et al.*, 1993). Clozapine is thus not recommended as a course of treatment unless previous administration of two other antipsychotics has failed (Zhang and Malhotra, 2013b).

The potentially detrimental side effects of antipsychotics significantly worsen compliance, lead to treatment discontinuation, and inhibit positive outcomes, necessitating the improvement of treatment strategies (Brandl *et al.*, 2014). To achieve mental well-being and ensure sustained quality of life for schizophrenia patients, these adverse reactions must be better understood and minimalised.

2.2.3. Treatment response

The goal of antipsychotic treatment is complete and sustained remission without relapse. However, much like other aspects of schizophrenia, treatment response is complex and heterogeneous, and this is rarely a reality (Robinson *et al.*, 2004). Although methods have not been standardised, the quantitative measurement of treatment outcome is commonly achieved with scales that measure symptom severity (Leucht *et al.*, 2008). For instance, general improvement is determined by comparing baseline and post-treatment BPRS scores, and changes in individual symptom domains are investigated with pre- and post-treatment PANSS, SANS, and SAPS scores (Remington *et al.*, 2010). In 2005, the Remission in Schizophrenia Working Group agreed upon criteria to define remission in the disorder (Andreasen *et al.*, 2005). Schizophrenia remission is achieved when particular core symptoms, such as hallucinations and blunted affect, are absent or mild (that is, they do not affect functioning) for at least six months. These criteria are unambiguous absolutes, as opposed to scale-specific degrees of symptom improvement, making them amenable to cross-study comparison (Emsley *et al.*, 2011).

There are several predictors of treatment outcome in schizophrenia. The most significant of these is the duration of untreated psychosis (DUP), which has an inverse relationship with positive outcome (Jeppesen *et al.*, 2008). Indeed, individuals experiencing their first episode of psychosis show 57-67% better response than those in more advanced stages of the disorder, highlighting the importance of early intervention (Emsley *et al.*, 2013). Another major influence on treatment efficacy is adherence to medication, with non-adherers five times more likely to relapse than adherent patients (Robinson *et al.*, 2004). This problem has largely been combatted by the replacement of oral administration with long-acting injectables (LAI; Nasrallah, 2007). Lastly, early response and nonresponse have been shown as reliable clinical markers for longer term outcome, with response at two weeks predictive of positive outcomes, and nonresponse indicative of treatment-refractoriness (Kinon *et al.*, 2010; Case *et al.*, 2011).

Unfortunately, remission is not achieved by the vast majority of patients, and approximately 50% of individuals show minimal to no response to antipsychotics (Lohoff and Ferraro, 2010). This is represented in Figure 2.5, which summarises the balance of good and poor outcomes in 18 independent studies on antipsychotics. All patients were experiencing their first episode of psychosis when recruited, and were monitored for more than one year post-treatment (van Os and Kapur, 2009). Nonresponse or treatment-refractoriness can be defined as a lack of improvement in symptoms after treatment with two different antipsychotics for at least six weeks each (Suzuki *et al.*, 2012). In these cases, clozapine is the go-to antipsychotic and has shown effective improvement in nonresponsive patients (Chowdhury *et al.*, 2011).

Considering the diverse scope of treatment outcomes, there is much to be discovered with regards to the workings of schizophrenia and antipsychotics. The heterogeneous clinical presentation of the disorder, high percentage of nonresponders, and severe ADR profiles of antipsychotics preclude the option of a standardised, one-for-all treatment design. Currently, genetic research into schizophrenia and antipsychotics is the starting point for developing individualised treatment and improved outcomes.

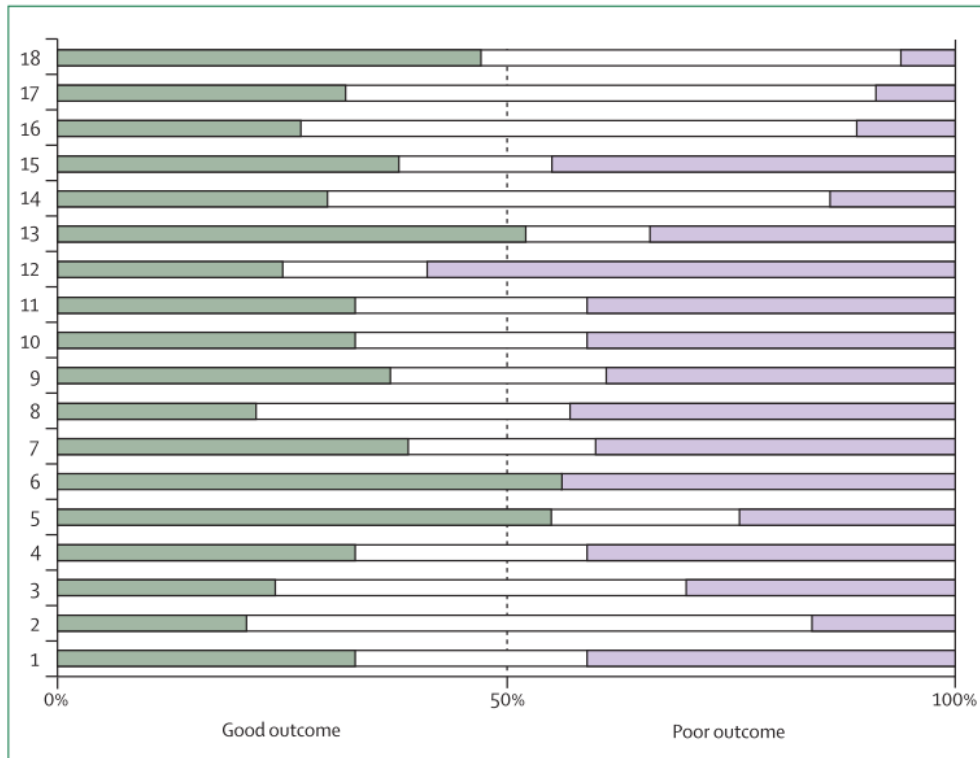


Figure 2.5: 18 studies demonstrating balanced outcomes of good vs. poor response to antipsychotic treatment (van Os and Kapur, 2009). *Reprinted with permission from Elsevier.*

2.3. Antipsychotic pharmacogenomics

2.3.1. Background

The term pharmacogenetics was created by Vogel in 1959 to explain the interaction between genetic differences on the range of treatment outcomes observed between individuals. Pharmacogenomics takes this a level further, by encapsulating differences across the entire genome that affect drug response. As for schizophrenia, antipsychotic response is considered to be a complex, multifactorial trait with a strong genetic basis (de Leon, 2009). Twin and family studies have demonstrated the high heritability of treatment response, including ADRs, and it is hypothesised that the genetic component of this heterogeneous phenotype is brought about by multiple variants of small effect across the genome (Arranz and de Leon, 2007; Sun *et al.*, 2012).

Researchers first investigated genetic predictors of schizophrenia treatment efficacy in the early 1990s, and many candidate pharmacogene studies have been performed since then (Zhang and Malhotra, 2013b). The roles of these genes in treatment response can be divided into two classes, namely pharmacodynamics and pharmacokinetics. The former refers to the interaction between a drug, transporters, and its target molecule(s), whilst the latter involves the absorption, distribution, and excretion of a drug (Zandi and Judy, 2010).

With regards to pharmacodynamics, considerable research has been performed on variation within dopamine receptor genes following the establishment of chlorpromazine's antidopaminergic action. Several polymorphisms in dopamine receptor genes, namely *DRD2*, *DRD3*, and *DRD4*, have shown associations with the extent of treatment efficacy and occurrence of ADRs in independent studies, the results of which are summarised by Arranz and Munro (2011). Furthermore, serotonin has been the secondary focus of pharmacodynamic studies. Alterations in the serotonergic system have been shown to play a role in both cognitive and negative symptoms of schizophrenia (Blanc *et al.*, 2010). As previously stated, this system is targeted by SGAs, and polymorphisms in both serotonin receptors and transporters have been implicated in treatment outcome and the extent of metabolic side effects (Blanc *et al.*, 2010).

In addition to neurotransmitter systems, numerous studies have been performed on the pharmacokinetics of antipsychotics, with a focus on drug-metabolising enzymes (DMEs) such as the Cytochrome P450 (CYP) family. *CYP2D6* codes for an enzyme essential for the majority of FGA metabolism (Lohoff and Ferraro, 2010), and is also highly polymorphic, with over 80 alleles having been identified (Rieder, 2014). This variation results in extreme individual differences, ranging from poor to ultra-rapid metabolism of drugs. Poor metabolisers of antipsychotics are at risk for developing drug toxicity and ADRs, whilst ultra-rapid metabolisers receive insufficient doses (Lohoff and Ferraro, 2010). In addition, *CYP1A2* is important for antipsychotic metabolism, and variation in this gene results in decreased enzyme activity (Murayama *et al.*, 2004). Other *CYP* polymorphisms have also been associated with variable treatment outcomes, such as those in *CYP3A4* and *CYP3A5* (Zandi and Judy, 2010). These studies have provided insight into the potential mechanisms of antipsychotics, but given the limited treatment success of drugs for the disorder, the candidate gene method has made way for more advanced, hypothesis-free approaches.

2.3.2. Genome-wide association studies

Unfortunately, the progress seen in schizophrenia susceptibility GWAS (2.1.4) is not matched by antipsychotic response GWAS. Only a handful of genome-wide studies have been conducted on the treatment response of schizophrenia, with the majority conducted in less than a thousand individuals per study (Alkelai *et al.*, 2009; Lavedan *et al.*, 2009). These GWAS are included in the National Human Genome Research Institute (NHGRI) GWAS Catalog, a database of all SNPs that have reached genome-wide significance ($P \leq 5 \times 10^{-8}$) for associations with one of 17 complex traits, including general drug response (Welter *et al.*, 2013). A diagrammatical layout of the Catalog is shown in Figure 2.6.

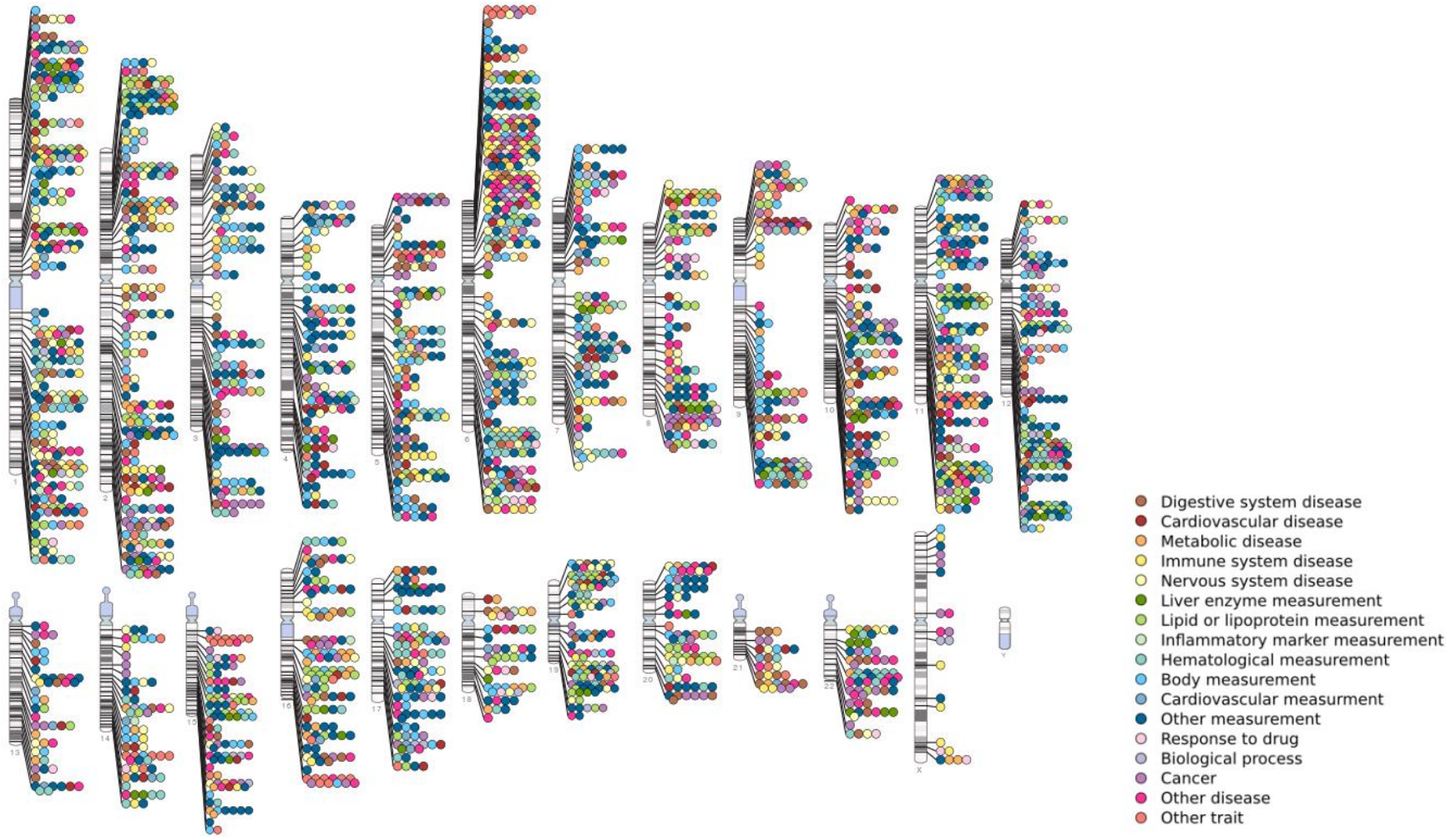


Figure 2.6: The NHGRI GWAS Catalog: 17 traits, including drug response, with significantly associated SNPs ($P \leq 5 \times 10^{-8}$) across the genome, as of December 2013 (www.genome.gov/gwastudies/). “Response to drug” includes GWAS on antipsychotic drug response.

Collecting samples of suitable size with sufficient statistical power for a GWAS is difficult for a complex phenotype such as antipsychotic response. This is likely because a large-scale study of schizophrenia treatment outcome ideally involves strict and replicable response criteria, as well as a clinically well-defined cohort (Meltzer, 2013). The gap between current pharmacogenomics research and clinical utility is vast. Zhang and Malhotra (2013b) state that “to date, there is no randomised clinical trial that provides solid support for using genetic testing to guide drug treatment in psychiatry”. However, antipsychotic pharmacogenomic GWAS do provide us with potential leads for new hypotheses on the mechanisms of both the drugs and the disorder, and the findings of these GWAS will be analysed in the current study.

2.4. Functional effects of genetic variation

2.4.1. Background

Focus in genetic research has shifted from identifying genes to determining their functions and revealing the biology that links genotype to phenotype (Auerbach *et al.*, 2013). With so-called “next-generation” approaches such as whole-genome sequencing and GWAS, there is a wealth of data available on genetic variation, which can provide us with clues to gene function, disease aetiology, and the manifestation of complex traits. The current challenge is to interpret the plethora of significant results in order to gain understanding of their mechanisms (Girard *et al.*, 2012).

Nonsynonymous variants in protein-coding regions have been widely studied. Predictive tools such as Polymorphism Phenotyping (PolyPhen-2; Adzhubei *et al.*, 2010) and Sorting Intolerant from Tolerant (SIFT; Kumar *et al.*, 2009) are commonly used to assess the functional impact of a SNP based on resulting amino acid changes. There are several databases of coding variants that have been linked to disease, including Online Mendelian Inheritance in Man (OMIM; McKusick, 1998) and The Human Gene Mutation Database (HGMD; Stenson *et al.*, 2009). With regards to pharmacogenomics, all variants that have been linked to drug response traits are curated in the PharmGKB database (Thorn *et al.*, 2010). This resource currently contains over 2000 genes, with information on associated drugs and clinical associations (<http://www.pharmgkb.org/>). Searching for the drug class “antipsychotics” returns 114 publications, the majority of which are candidate gene studies. These examples illustrate that coding variants have been studied extensively. This is due to their being amenable to functional analyses; that is they demonstrate a clear phenotypic alteration (Cooper and Shendure, 2011). However, noncoding variants have recently come into focus and have important implications for complex traits such as antipsychotic response.

2.4.2. Noncoding variation

Even though studies on coding variation have revealed harmful disease-associated changes at the protein level, approximately 98% of the genome contains untranslated DNA and thus the majority of genomic variation. Some noncoding regions are transcribed into regulatory noncoding RNAs, whilst others act as sequence-specific binding sites for regulatory machinery that subsequently affect gene expression (Cooper and Shendure, 2011).

Gene regulation is a multilevel process involving *cis*-elements, namely promoters, silencers, and enhancers, and *trans*-elements such as transcription factors that bind to *cis*-elements (Georgitsi *et al.*, 2011). Illustrating the complexity of regulation, transcription factors are known to bind to thousands of different downstream target sites (Cooper and Shendure, 2011). Furthermore, epigenetic mechanisms and environmental factors also influence gene expression (Pastinen, 2010).

Polymorphisms in noncoding regions of the genome may contribute substantially to complex trait phenotypes. It is hypothesised that perturbations in regulatory pathways may have widespread effects and be significant in schizophrenia development, severity, and antipsychotic response (Arranz and de Leon, 2007). Substantiating this, the majority of common SNPs associated with disease lie in noncoding regions (Freedman *et al.*, 2011). Regulatory SNPs (rSNPs) affect the phenotype by modifying gene expression, and may produce larger effects than coding mutations. Georgitsi and colleagues (2011) propose that rSNPs likely create greater alterations in the level of protein product, since coding SNPs do not necessarily affect efficient transcription and translation.

Despite the hypothesised functional consequences of rSNPs, there is a shortage of studies demonstrating their effect. This is due, in part, to the difficulty of studying these regions. For example, tools such as PolyPhen-2 are based on evolutionary constraint analysis, an approach that is feasible for protein-coding regions because they are highly conserved. However, regulatory regions undergo many more changes over time and are not amenable to this type of analysis (Schmidt *et al.*, 2010). Furthermore, the complexity of gene regulation and lack of information on the mechanisms of this multifactorial system have made characterising rSNPs difficult (Drögemöller *et al.*, 2014a; Ritchie *et al.*, 2014).

Employing candidate gene approaches, pharmacogenetics has investigated a handful of regulatory regions flanking genes coding for DMEs, drug transporters and targets. Selected examples of rSNPs for treatment outcomes are listed in Table 2.2. Although these studies are hypothesis-bound, they suggest the importance of rSNPs in altering the response

phenotype via expression changes in well-studied genes. Indeed, patients are commonly genotyped for the *VKORC1* promoter variant prior to warfarin administration, and dosage is tailored accordingly to avoid warfarin-resistance (International Warfarin Pharmacogenetics Consortium, 2009). There are many more pharmacogenes than those listed in the table, but a scarcity of studies on related regulatory variants (Georgitsi *et al.*, 2011).

Table 2.2: Selected rSNPs associated with changes in expression of pharmacogenes.

Gene	Variation	Disease and/ or effect	References
<i>UGT1A1</i>	c.-3279T>G	<i>Gilbert's syndrome</i> Decreased bilirubinemia in G allele carriers	Sugatani <i>et al.</i> , 2002
<i>CYP2C19</i>	c.-806C>T c.-3402C>T	<i>Psychiatric disorders</i> Poorer treatment outcomes in T homozygotes	Sim <i>et al.</i> , 2006 Li-Wan-Po <i>et al.</i> , 2010
<i>CYP2D6</i>	c.-1584C>G	<i>Psychiatric disorders</i> Ultrarapid thioridazine metabolism in G allele carriers	Zanger <i>et al.</i> , 2001 Dorado <i>et al.</i> , 2009
<i>CYP3A4</i>	c.522-191C>T	<i>Cardiovascular disease</i> T allele carriers require lower statin doses	Wang <i>et al.</i> , 2010
<i>SLCO1B1</i>	g.-11187G>A in linkage with c.521T>C(*5)	<i>Hypercholesterolemia</i> Reduced pravastatin transport activity and reduced cholesterol concentration reduction	Niemi <i>et al.</i> , 2004 Kameyama <i>et al.</i> , 2005
<i>VKORC1</i>	c.-1639G>A	<i>Warfarin-resistance</i> G allele carriers require higher dose	Yuan <i>et al.</i> , 2005 Flockhart <i>et al.</i> , 2008

c. = mutation in coding DNA; g. = mutation in genomic DNA.

The development of hypothesis-free, genome-wide approaches allows for the exploration of noncoding variation within the genome. In fact, the majority of observed GWAS “hits” are located in noncoding regions (Hindorff *et al.*, 2009; Adkins *et al.*, 2011). Unfortunately, characterisation of these regions in previous GWAS is rare, because most studies tend to focus solely on the functionality of adjacent genes, without investigating the potential role of the polymorphism in regulatory networks (Schaub *et al.*, 2012; Ritchie *et al.*, 2014). Restricting focus to the closest gene creates problems, since a significant polymorphism may only be a proxy for the causal SNP due to linkage disequilibrium (LD). Research has shown that significant variants can be in perfect LD with SNPs that are hundreds of kilobases away, sometimes outside of the range accounted for by the GWAS tag SNP (Schaub *et al.*, 2012). Alternatively, a nearby SNP in perfect LD with the associated variant may be ignored in the interpretation phase of GWAS, because it falls within a noncoding region.

Recently, characterisation of noncoding variation has focused on expression quantitative trait loci (eQTLs). eQTLs have been identified by combining whole-genome approaches such as sequencing and GWAS with quantification of genome-wide expression levels. This

allows for a link between variant identification and functional classification (Cookson *et al.*, 2009). These noncoding loci are associated with the expression of a particular gene that can be either proximally or distally located (Pastinen, 2010). Recently, Qiu *et al.* (2014) used over 300 expression microarrays to assess gene expression levels in treated and untreated childhood asthma. The results constituted over 2000 drug-response eQTLs associated with several hundred genes; many compelling new targets for asthma treatment research. Another study conducted by Mamdani and colleagues (2013) investigated peripheral gene expression in antipsychotic treatment response, and identified 22 differentially expressed genes between responders and nonresponders. Importantly, many of these genes contain eQTLs that could be used as biomarkers for treatment response in future work. These studies serve as examples of the benefit of eQTL studies in pharmacogenomics.

2.4.3. Recent bioinformatic developments

In the wake of newer genomic techniques generating an abundance of genotype data, result interpretation and functional analyses have fallen behind. The mass of results generated has little clinical utility and our improvement in understanding complex traits is minimal. However, the development of bioinformatic tools to examine the functional implications of genomic variation is rapidly breaking through this research bottleneck.

The first of its kind, the Encyclopedia of DNA Elements (ENCODE) sought to characterise and make publicly available all the functional elements of the genome. A functional element is defined by its creation of a reproducible biochemical signal (The ENCODE Project Consortium, 2007). This definition includes coding and noncoding DNA, noncoding RNAs, and *cis*-regulatory elements. The concept of functional elements is illustrated in Figure 2.7. In order to characterise genomic regions, several experimental approaches were used, including chromatin immunoprecipitation and sequencing (ChIP-seq), to identify transcription factor binding sites (TFBS) across the genome. Additionally, the binding of regulatory factors to *cis*-elements produces changes in chromatin states, which were identified by deoxyribonuclease I (DNase I) hypersensitive site (DHS) mapping (Maurano *et al.*, 2013).

The results generated by ENCODE were overwhelming: 1640 data sets performed on 147 different cell types, with the controversial conclusion that 80% of the genome is functional, much of it previously believed to be “junk” DNA (The ENCODE Project Consortium, 2011). Interestingly, ENCODE has revealed that rSNPs are significantly enriched for GWAS associations (Schaub *et al.*, 2012). Conversely, noncoding variants associated with GWAS were found to be concentrated in regions implicated in regulation (Maurano *et al.*, 2013).

The functional annotation of the genome has mapped out regulatory regions and provided links between regulatory elements and gene expression. This plethora of findings enables researchers to analyse functional implications of noncoding variation, to predict whether they significantly alter a regulatory element. For example, in a recent schizophrenia susceptibility GWAS by the PGC, the authors used ENCODE to analyse the overlap of significant variants with DHSs, in order to assess their regulatory potential (Ripke *et al.*, 2013). The ENCODE findings are encouraging for future antipsychotic treatment response studies. With the vast amounts of new information at our fingertips, progress in the development of new drugs, as well as improvement in the use of existing ones, becomes a possibility.

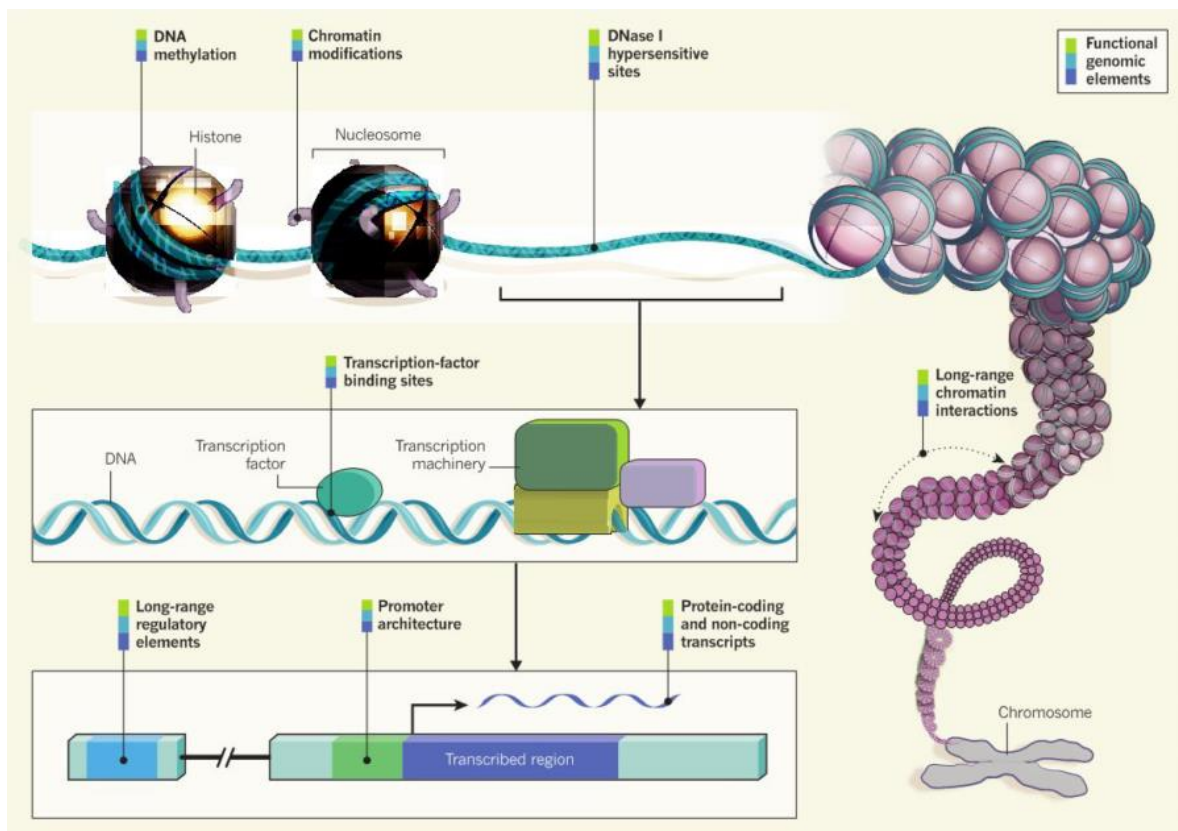


Figure 2.7: Illustration of the various types of functional elements within the genome defined by ENCODE (Ecker *et al.*, 2012). Reprinted with permission from Nature Publishing Group.

Subsequent to the ENCODE project, several bioinformatic tools were created to pinpoint the biological implications of rSNPs; a task that was previously impossible on a large scale (Cooper and Shendure, 2011). Several examples are listed here. Firstly, RegulomeDB is a database that incorporates ENCODE datasets, computational predictions, a large amount of eQTLs, and other published literature to predict the effect that a single base pair change can have on the binding of regulatory elements across the genome (Boyle *et al.*, 2012). This tool

has been widely used to analyse the impact of noncoding SNPs on gene regulation (Chung *et al.*, 2013; Juraeva *et al.*, 2014).

Another bioinformatic tool, Transcription factor Affinity Prediction (TRAP), is utilised to assess the impact of SNPs on known transcription factor motifs (Manke *et al.*, 2010). An application specifically designed to predict the impact of single nucleotide changes, sTRAP, compares “wild-type” and “mutant” sequences, i.e. potential rSNPs, and evaluates subsequent changes in affinity for transcription factors (TFs) to known TFBS motifs.

rSNPBase is a resource of curated rSNPs identified by experimental analysis including ENCODE. Additionally, rSNPs uncovered by eQTL studies and computational predictions are included (Guo *et al.*, 2014). The data is used to assess the functionality of noncoding variants, as well as predict which downstream genes are affected in which tissue types. This tool would be useful for discovering regulators of pharmacogenes, novel candidate pharmacogenes, and their corresponding expression levels in relevant tissues such as the brain and liver.

Also analysing genome-wide expression levels, the functional annotation of the mammalian genome 5 (FANTOM5; The FANTOM Consortium *et al.*, 2014) is a database of mammalian gene expression patterns across different tissues. This catalogue provides functional annotation of rSNPs and cell-specific transcriptome profiles, and has information available for at least one promoter for 95% of protein-coding genes.

These resources allow scientists to broaden their focus to whole-genome functionality. Inclusion of all possible candidate SNPs via a hypothesis-free approach, coupled with in-depth functional analyses, increases the likelihood of uncovering true causal variants. If a noncoding causal variant is elucidated, downstream gene targets can be predicted, and their expression levels in relevant tissues can be determined. This could improve our understanding of the complex biological mechanisms behind antipsychotic response. Additionally, such biological insights could lead us to biomarkers that improve diagnosis and treatment of schizophrenia.

2.5. The South African context

As discussed in chapter one, South Africa is home to unique and genetically diverse population groups. Pharmacogenetic studies on South African individuals have studied this variation with regards to drug response, contributing to the narrowing of the “10/90” research

gap between high- and low-to-middle income countries. For example, the population-specific diversity of the *CYP* alleles was illustrated by Gaedigk and Coetsee (2008) in a study of 99 SAC individuals. The authors uncovered two novel *CYP2D6* alleles, and observed frequencies vastly different to previously characterised variation in individuals of European descent. Both of these findings point to diminished *CYP* enzyme activity, and thus the need for ethnicity-specific treatment regimens. *CYP2D6* is an important metaboliser of antipsychotics, suggesting that these results have implications for schizophrenia treatment as well as other drug responses (Zanger *et al.*, 2004). Mitchell and associates (2011) also identified 26 novel alleles in the *CYP2C9* gene in a black South African cohort. Additionally, this study investigated variation in the well-studied *VKORC1* gene, associated with warfarin dosage. Variation in these two genes – as well as minor environmental covariates accounted for in the study – were found to contribute to approximately 45% of the heterogeneity in warfarin dosage. In a larger-scale study, Ikediobi and colleagues (2011) genotyped over 200 SNPs across 12 genes that have been previously associated with antiretroviral treatment response in two genetically unique South African population groups. Significant differences in *CYP* allele frequencies were seen between the two groups, illustrating the diversity of South African ancestry and the heterogeneity of treatment response (Ikediobi *et al.*, 2011; Warnich *et al.*, 2011).

With specific reference to antipsychotic pharmacogenomics, a novel approach applied exome sequencing to 11 South African first episode schizophrenia (FES) patients, followed by variant prioritisation and genotyping in a larger FES and Xhosa cohort (Drögemöller *et al.*, 2014b). Several loss-of-function variants were identified, the majority previously unidentified or at very low frequencies in Asian and European population groups. Once again, this study highlights the uniqueness of South African genomes and the need for increased research in this field.

Unfortunately, the pharmacogenomic studies in South Africa have mostly spanned a handful of candidate genes in relatively small sample sizes (Warnich *et al.*, 2011). Due to high incidence of diseases such as HIV and tuberculosis, the few GWAS that have been performed on Southern Africans have focused on these illnesses (Petrovski *et al.*, 2011; Chimusa *et al.*, 2014). Globally, GWAS are performed at a ratio of approximately 10:1 European ancestry vs. all other ancestry groups combined (Need *et al.*, 2009). To date, only seven GWAS have been performed exclusively on African individuals, with four others including some Africans, none of these focusing on schizophrenia treatment response (<http://www.genome.gov/gwastudies/>). This is problematic, since the allele frequency of a GWAS variant can vary up to 40-fold between population groups (Adeyemo and Rotimi,

2010), and patterns of linkage disequilibrium are vastly different between Caucasian and African genomes (Dalal *et al.*, 2010). Therefore, the progress made in developed countries in schizophrenia risk and treatment GWAS may not be applicable to South Africans.

These factors call for increased study of the genetic diversity of South African individuals, as well as psychiatric pharmacogenomic research applications, in the hope of improving and understanding treatment response in schizophrenia in a population-specific manner. Optimising treatment with the use of pharmacogenomics is particularly important in LMIC, since countries such as South Africa experience a greater health burden. Even though there is much research to be done, South Africa is considered the leader amongst developing countries in pharmacogenomics and biomarker research (Gupta *et al.*, 2014).

2.6. Overview of the current study

2.6.1. Aim and objectives

This study aims to explore the functional consequences of noncoding genetic variants that contribute to complex and heterogeneous antipsychotic treatment outcomes in schizophrenia, and subsequently investigate these findings in South African individuals.

The specific objectives of this study are as follows:

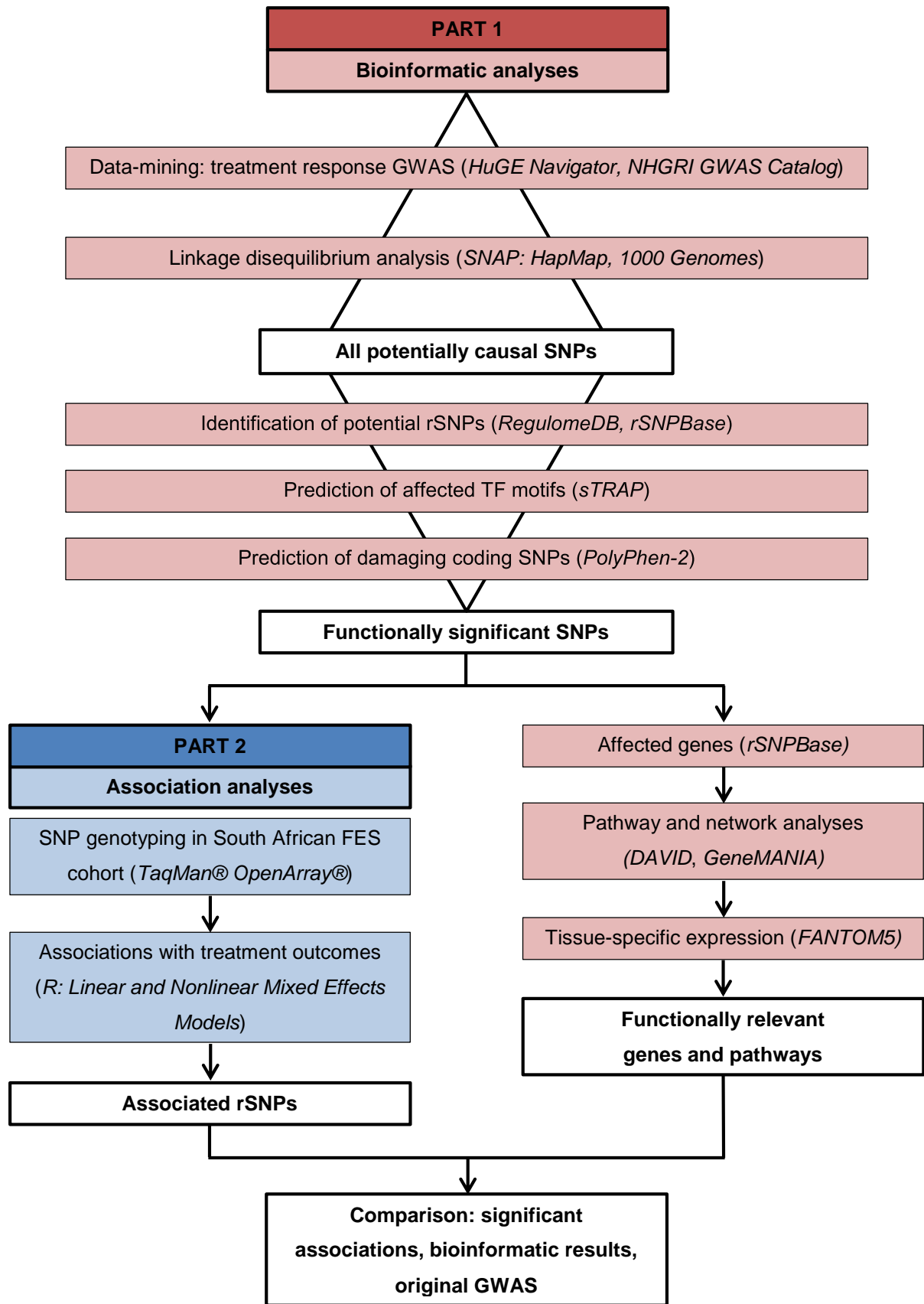
PART 1: Bioinformatic identification of potential regulatory variants associated with antipsychotic treatment response

- Survey the literature to identify GWAS SNPs that are significantly associated with antipsychotic treatment response, including ADRs, in schizophrenia.
- Make use of publicly available genetic variation data to determine the variants in LD with the associated SNPs.
- Employ recent publicly available data and bioinformatic tools to identify noncoding SNPs that potentially affect regulation.
- Identify affected genes and pathways and assess likelihood of their involvement in treatment response.

PART 2: Associations between predicted regulatory variants and antipsychotic treatment outcomes in a South African schizophrenia cohort

- Prioritise SNPs and genotype in a South African first episode schizophrenia cohort.
- Perform association analyses to determine which variants are associated with treatment outcomes within the cohort.
- Compare associations with regards to original GWAS SNPs.

2.6.2. Strategy



PART 1

Bioinformatic analyses

CHAPTER 3:

Bioinformatic identification of potential regulatory variants associated with antipsychotic treatment response

3.1. Summary

Advances in genetic data generation have increased exponentially in recent years. For example, results of the ENCODE project provide masses of novel information on noncoding DNA and gene regulation. To interpret the multitude of results, bioinformatic resources are constantly being developed and improved upon. These tools allow in-depth interpretation, including uncovering the functional implications of noncoding variants associated with complex traits. One such trait is antipsychotic treatment response. Antipsychotics are the most effective treatment for schizophrenia, a debilitating psychiatric disorder, but up to half of patients respond poorly to these drugs and can develop adverse drug reactions.

To improve our limited understanding of the biological mechanisms underpinning schizophrenia treatment outcomes, a novel bioinformatics pipeline was applied in this study. Previously significant GWAS variants ($P \leq 5 \times 10^{-7}$) were mined, and publically available population data was used to find SNPs in LD ($r^2 \geq 0.8$) with these variants. This approach allowed for hypothesis-free evaluation of genome-wide variants. Subsequently, tools making use of ENCODE data, namely RegulomeDB and rSNPBase, predicted the regulatory impact of the variants and their affected gene targets, including eQTLs. Pathway and network analyses were performed with DAVID and GeneMANIA respectively, and tissue-specific expression of the affected genes was assessed by FANTOM5.

Despite few GWAS and poorly-characterised cohorts, several regulatory effects were identified. The most important results of this study were that i) treatment-refractory schizophrenia was the most common trait significant for regulation, with 16 rSNPs on 4q24 affecting *NFKB1* expression, and ii) there was extensive overlap with regions and functions that had been previously implicated in schizophrenia risk, particularly with regards to immune dysregulation. *NFKB1*, which also plays an important role in immune functioning and has been previously implicated in schizophrenia, was affected by almost half of the identified rSNPs. The significance of *NFKB1* was confirmed with pathway and network analyses, which illustrated interactions with other genes for three of four predicted pathways. Lastly, gene expression analyses showed that four of the top 10 affected genes were most upregulated in brain tissues. This study provides evidence for the overlap between schizophrenia risk and treatment response. Additionally, the importance of well-

characterised cohorts for clinical and genetic analyses is clear. *NFKB1* and other associated genes should be further studied in different, population-specific, and replicative cohorts in order to validate potential regulatory biomarkers of treatment response.

3.2. Introduction

In the last decade there have been major advancements in elucidating the genetics of complex traits and diseases. GWAS enable the simultaneous analyses of hundreds of thousands to millions of variants across the genome. This approach is popular, shown by the thousands of GWAS recorded in the NHGRI GWAS Catalog (Welter *et al.*, 2013). Despite the abundance of data, the biological interpretation of genetic signals associated with particular traits is mostly insufficient, and subsequently misguided. The majority of significant GWAS variants lie in noncoding regions. Even so, biological interpretations tend to assess SNPs in terms of the function of the closest gene, even if the SNP is intergenic and hundreds of kilobases away (Ritchie *et al.*, 2014). This approach restricts interpretation, particularly since the closest gene may not incorporate all variants within the tag SNP's haplotype, and LD can vary extensively between population groups (Christoforou *et al.*, 2012).

Traditionally, coding SNPs have proved far more amenable to functional analyses, which means that they are often prioritised post-GWAS for further study, whilst noncoding variants are ignored (Cooper and Shendure, 2011). Recently, we have learnt a great deal more about noncoding regions with the results of the ENCODE project, which sought to characterise all the functional elements of the genome, including regulatory factors (The ENCODE Project Consortium, 2011). With this knowledge, it has been revealed that noncoding SNPs implicated in regulation, or rSNPs, are enriched for GWAS associations, highlighting the importance of analysing these regions for implications in disease (Schaub *et al.*, 2012; Maurano *et al.*, 2013). Indeed, rSNPs are potentially more damaging than coding SNPs, considering the wider range of expression dysregulation associated with transcription and translation candidates (Georgitsi *et al.*, 2011).

Many bioinformatic tools that make use of the abundance of ENCODE data have recently been developed. For example, RegulomeDB (Boyle *et al.*, 2012) and rSNPBase (Guo *et al.*, 2014) both assess the regulatory potential of a SNP, the former looking at eQTL evidence and proximal regulation, and the latter assessing proximal, distal and post-transcriptional functioning and predicting downstream gene targets.

Schizophrenia, an example of a widely studied heterogeneous disease, is a chronic and debilitating psychiatric disorder. Antipsychotics are the main treatment for this disease, although treatment proves ineffective for roughly a third of patients (Liou *et al.*, 2012). Despite the fact that genetic variation has shown to substantially contribute to antipsychotic treatment response, the mechanisms involved are not well understood. Consequently, in the last five years, GWAS have been applied to schizophrenia treatment response in an attempt to shed light on the genetics of poor treatment response and adverse drug reactions.

The purpose of this study was to analyse previous antipsychotic response GWAS and investigate functionality with the use of these recent bioinformatic advancements. By incorporating LD variants from HapMap (The International HapMap Consortium, 2003) and 1000 Genomes (The 1000 Genomes Project Consortium, 2010) populations, all the potential “causal” SNPs to date were accounted for. These variants were then investigated with the use of experimentally validated data as well as predictive tools to isolate rSNPs and their gene targets. Subsequently, pathways, networks, and tissue-specific expression of these genes were assessed. To our knowledge, this is the first study of its kind to investigate antipsychotic response GWAS with regards to genomic regulation, and the use of this bioinformatics pipeline has the potential to improve our understanding of the biology of treatment response in schizophrenia.

3.3. Materials and methods

Refer to Electronic Sources (p. 107) for dates of access to online tools.

3.3.1. Data-mining

The literature was mined in order to identify all variants from GWAS that have been significantly associated with antipsychotic response in schizophrenia, including ADRs. This was accomplished with the use of a database of complex disease GWAS, the human genome epidemiology (HuGE) Navigator GWAS Integrator version 2.0 (Yu *et al.*, 2008), available at <http://hugenavigator.net/HuGENavigator/home.do>. The search terms “antipsychotic”, “schizophrenia”, and “adverse drug reaction” were used successively to identify all relevant GWAS. The corresponding articles were accessed in PubMed via HuGE Navigator and further investigated. HuGE Navigator is updated regularly, but a survey of the NHGRI GWAS Catalog (<http://www.genome.gov/gwastudies/>) and a manual literature search of PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>) were performed with the same search terms to account for any studies that may not have been included on HuGE Navigator.

Due to multiple testing, it is generally accepted that the genome-wide significance threshold should be $P \leq 5 \times 10^{-8}$ (Sham and Purcell, 2014). Using this cut-off point, less than 10 SNPs from four GWAS qualified for further analysis (highlighted in Table 3.3), over half of which related to antipsychotic-induced metabolic side effects (Adkins *et al.*, 2011; Malhotra *et al.*, 2012). Due to the restrictive nature of these results, the threshold for inclusion of SNPs was raised to $P \leq 5 \times 10^{-7}$ in order to increase the amount of variants, independent studies, and response phenotypes. Kingsmore and colleagues (2008) refer to this value as a “respected threshold” in a review of GWAS study designs.

Additionally, the SNPs were investigated with SeattleSeq Annotation 137 version 8.07 (<http://snp.gs.washington.edu/SeattleSeqAnnotation137/>) to confirm their position with respect to the genes listed in the nine GWAS, since there is a tendency to relate the function of the closest gene to the effect of the SNP, without considering the variant’s potential role in other pathways. SeattleSeq acquires gene information from the Human Genome Organisation (HUGO) Gene Nomenclature Committee (HGNC; <http://www.genenames.org/>).

3.3.2. Variants in linkage disequilibrium

To find SNPs in high LD ($r^2 \geq 0.8$; Carlson *et al.*, 2004) with the significant GWAS variants, SNP Annotation and Proxy search (SNAP) version 2.2 was employed (<http://www.broadinstitute.org/mpg/snap/>). This tool conveniently combines two large-scale and widely used repositories of human genetic variation, namely HapMap (The International HapMap Consortium, 2003) and 1000 Genomes (The 1000 Genomes Project Consortium, 2010). SNAP allows one to choose between four data sets: 1) 1000 Genomes Pilot 1, 2) HapMap phase II release 21, 3) HapMap phase II release 22, and 4) HapMap phase 3 release 2. For each set, one can specify which population group to analyse. 1000 Genomes and HapMap population groups available on SNAP are indicated in Table 3.1.

Different population groups were analysed depending on the ancestral make-up of the patient samples in the GWAS. For example, many of the identified GWAS studied SNPs within the CATIE cohort (discussed in 2.2.1), which consisted of 57% Caucasian, 29% African American and 14% “other” individuals (McEvoy *et al.*, 2005). Therefore, for CATIE, all population groups on SNAP were analysed to account for LD structures within these diverse ancestry groups. In another GWAS (Liou *et al.*, 2012), all individuals under study were of Han Chinese descent, therefore only the CHBJPT and JPT+CHB+CHD population options were used for analysis of these SNPs.

Table 3.1: Population groups on SNAP (<http://www.broadinstitute.org/mpg/snap/>).

Population code	Description	Dataset
CEU	Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain (CEPH) collection	1, 2, 3, 4
YRI	Yoruba in Ibadan, Nigeria	1, 2, 3, 4
CHBJPT	Han Chinese in Beijing, China, and Japanese in Tokyo, Japan	1, 2, 3, 4
ASW	African ancestry in south west America	4
CHD	Chinese in metropolitan Denver, Colorado	4
GIH	Gujarati Indians in Houston, Texas	4
LWK	Luhya in Webuye, Kenya	4
MEX	Mexican ancestry in California, Los Angeles	4
MKK	Maasai in Kinyawa, Kenya	4
TSI	Toscans in Italy	4
CEU+TSI	Combined panel of CEU and TSI	4
JPT+CHB+CHD	Combined panel of JPT, CHB and CHD	4

1 = 1000 Genomes Pilot 1; 2 = HapMap phase II release 21; 3 = HapMap phase II release 22; 4 = HapMap phase 3 release 2.

Separate analyses for each dataset for each relevant population were performed. Subsequently, the results were manually combined into the appropriate population supergroups as defined in this study (Caucasian: CEU, CEU+TSI; African: YRI, ASW, LWK, MKK; and other: CHBJPT, CHD, GIH, MEX, JPT+CHB+CHD). For each GWAS, only SNPs in common between the supergroups were included for further analyses. Additionally, SNPs were excluded if they occurred in either a HapMap or 1000 Genomes dataset, but not the other. The SNPs were, therefore, filtered to include only those that were i) in LD with a GWAS SNP with an r^2 equal to or greater than 0.8, and ii) present in *all* relevant population supergroups in *both* 1000 Genomes and HapMap datasets (1 and 2 or 3 or 4).

3.3.3. RegulomeDB analysis

Original GWAS variants and SNPs in LD were analysed with several bioinformatic resources. RegulomeDB (<http://regulome.stanford.edu/>) is a database that incorporates ENCODE datasets, computational predictions, a large amount of eQTLs, and other published literature to predict the effect that a single base pair change can have on the binding of regulatory elements to DNA (Boyle *et al.*, 2012). This tool has been widely used to analyse the impact of noncoding SNPs on gene regulation (Chung *et al.*, 2013; Juraeva *et al.*, 2014).

RegulomeDB annotates intergenic SNPs within regulatory elements, and predicts the degree to which a SNP will interfere with binding and downstream regulatory processes. Regulatory elements include promoters, enhancers, and TFBS. Predictions are made using a heuristic scoring system based on the amount of evidence available for a particular SNP, as illustrated in Table 3.2.

Table 3.2: RegulomeDB scoring system, with category 1 being most significant and category 6 least significant (Boyle *et al.*, 2012).

Category	Description
	<i>Likely to affect binding and linked to expression of a gene target</i>
1a	eQTL + TF binding + matched TF motif + matched DNase footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding/DNase peak
	<i>Likely to affect binding</i>
2a	TF binding + matched TF motif + matched DNase footprint + DNase peak
2b	TF binding + any motif + DNase footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak
	<i>Less likely to affect binding</i>
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif
	<i>Minimal binding evidence</i>
4	TF binding + DNase peak
5	TF binding or DNase peak
6	Motif hit

According to Boyle *et al.* (2012), eQTLs classify a SNP as being highly significant, since eQTLs display direct experimental evidence for altering expression levels of a particular gene. Therefore, SNPs in category 1 are more significant than any other evidence for regulation. Category 1 decreases in significance from subcategories “a” to “f”, with “a” having more annotations available than “b”, and so on. These annotations include experimental evidence for transcription factor (TF) binding, and computational predictions such as DNase footprinting. Category 2 is identical to category 1, with the exception of eQTL evidence. Category 3 consists of SNPs with less evidence of regulatory impact and categories 4-6 exhibit minimal evidence thereof (Boyle *et al.*, 2012).

The list of combined GWAS and LD SNPs was entered into RegulomeDB and SNPs were ranked according to the scoring system. Seven SNPs returned a server error. These rs numbers were entered into the Ensembl Genome Browser

(<http://www.ensembl.org/index.html>) to identify synonyms, and the new rs numbers were successfully analysed by RegulomeDB. SNPs in categories 1-3 are considered important for further investigation, and were prioritised for further study (Chung *et al.*, 2013).

3.3.4. rSNPBase analysis

The rSNPBase (<http://rsnp.psych.ac.cn/>) allows for the assessment of different types of regulation, including proximal, distal and post-transcriptional processes (Guo *et al.*, 2014). For further insight into the regulatory potential of antipsychotic response variants, the GWAS and LD SNPs were analysed with rSNPBase to identify potential rSNPs. Additionally, genes predicted to be affected by these rSNPs were retrieved.

3.3.5. Variants affecting binding motifs

Concurrent to the rSNPBase and RegulomeDB analyses, Transcription factor Affinity Prediction (TRAP) was utilised to assess the impact of the SNPs on known transcription factor motifs (Manke *et al.*, 2010). A tool specifically designed to predict the impact of single nucleotide changes, sTRAP (http://trap.molgen.mpg.de/cgi-bin/trap_two_seq_form.cgi), compares “wild-type” and “mutant” sequences, and evaluates subsequent changes in affinity for TFs to known TFBS motifs, attained from the Jaspas database (<http://jaspar.genereg.net/>). In this computational model, changes in affinity are calculated based on the log ratio of binding probability between “wild-type” and “mutant” sequences (Manke *et al.*, 2010).

The R package (R Development Core Team, 2010) of sTRAP, tRap, is freely available and was utilised to evaluate all SNPs simultaneously (<http://trap.molgen.mpg.de/cgi-bin/download.cgi>). A custom Unix script incorporating tRap was designed by Dr N. Ishaque of the German Cancer Research Center (Heidelberg, Germany; Script S1). These commands were performed in the Unix shell and the output was exported to a .txt file. Firstly, a list of all the SNPs were converted to a Browser Extensible Data (BED) file, for which a FASTA file was generated. Similarly, BED and FASTA files for the regions 15 base pairs up- and downstream of each variant were created. A Perl script was written by Dr Ishaque to combine the FASTA files (Script S2), in order to analyse the sequences with tRap.

The results were ranked by log ratio, r , with the largest positive or negative value indicating the most significant increase or decrease in binding affinity respectively, with reference to the “wild-type” sequence (Manke *et al.*, 2010). Significant motifs present in humans were obtained from Jaspas (<http://jaspar.genereg.net/>) from the TRAP motif identifiers.

To assess the uniqueness of results obtained for schizophrenia treatment response, the same workflow was applied to three other complex traits. HuGE Navigator was surveyed for drug response GWAS, and two traits with sufficient results were analysed further, namely response to antidepressants, and response to hepatitis C treatment. Response to antidepressants was selected because, as previously discussed, psychiatric disorders have shown biological overlap in susceptibility and treatment response. On the other hand, hepatitis C treatment response has not been shown to be related to antipsychotic response. Additionally, GWAS on variants associated with eye colour, a complex trait, were analysed. The purpose of this was to establish a “baseline” level of regulation within the genome and to determine if any motifs were enriched for drug response, and more specifically, response to antipsychotics. These “control” traits were evaluated using the methods outlined in 3.3.2 and then assessed with the customised tRap script (Script S1).

3.3.6. Nonsynonymous coding variants

As an additional analysis, SNPs in coding regions were assessed for effects on protein function. To achieve this, PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>) was used. This resource uses computational prediction to characterise nonsynonymous SNPs based on their predicted impact on protein-coding regions of the genome (Adzhubei *et al.*, 2010). PolyPhen-2 classifies SNPs as either “benign”, “possibly damaging”, or “probably damaging”. Additionally, a score is assigned that translates as the probability of the variant being damaging, i.e. a score closer to one indicates a more damaging SNP (Adzhubei *et al.*, 2010).

3.3.7. Affected genes and pathways

To assess the affected genes, functional clustering was performed with the use of pathway and network analyses. A list of genes was compiled containing i) eQTL targets from RegulomeDB and rSNPBase, ii) affected genes according to rSNPBase, and iii) genes containing any nonsynonymous variants classified by PolyPhen-2.

The Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.7 is a commonly used online pathway analysis tool available at <http://david.abcc.ncifcrf.gov/home.jsp> (Huang *et al.*, 2009). By entering the gene list into DAVID, the aim was to investigate whether the genes were common to any pathways, and what the biological significance of these pathways is. The Functional Annotation Clustering tool was used to identify pathways via several databases: the Biological Biochemical Image Database (BBID), the Kyoto Encyclopedia of Genes and Genomes (KEGG), BioCarta, Protein Analysis Through Evolutionary Relationships (PANTHER) and Reactome.

Clustered genes from DAVID were further explored with GeneMANIA (<http://www.genemania.org/>). This graphical interface forms connections between genes based on co-expression, cell or tissue co-localisation, genetic and physical interactions, and predicted functional relationships (Warde-Farley *et al.*, 2010). Relevant genes were grouped and visualised together based on functional and pathway classifications from DAVID. Subsequently, these annotations were assessed for their likelihood to be of relevance to antipsychotic treatment response.

3.3.8. Tissue-specific gene expression

Affected genes were assessed with regards to tissue-specific activation. Firstly, genes were ranked according to the number of rSNPs predicted to affect them. Subsequently, to determine the gene's potential involvement in antipsychotic response mechanisms, expression levels were identified in healthy brain and liver cell lines via the Semantic catalogue of Samples, Transcription initiation And Regulators (SSTAR) on the FANTOM5 database (<http://fantom.gsc.riken.jp/5/sstar>). This resource uses Cap Analysis of Gene Expression (CAGE) sequencing to identify cDNA and subsequently map it to a particular transcription start site. The number of sequenced fragments, or "tags", correlates with the level of gene expression in that particular tissue (Kodzius *et al.*, 2006). Expression is measured in tags per million (TPM), where a gene is considered to be "switched on" if it shows at least 10 TPM (The FANTOM Consortium *et al.*, 2014). Cell lines relating to brain and liver tissue on FANTOM5 are listed in Box S1. Each brain and liver cell line was analysed for each gene, and the cell line with the highest expression levels was also recorded for comparison.

3.4. Results

3.4.1. Antipsychotic response GWAS

The HuGE Navigator GWAS Integrator and the NHGRI Catalog were mined to identify previous GWAS on antipsychotic pharmacogenomics. Nine studies with SNPs reaching genome-wide significance ($P \leq 5 \times 10^{-7}$) were identified. All but one GWAS was found by both databases. The most recent study, by Clark *et al.* (2013), was only listed in the NHGRI Catalog, most likely because it is more regularly updated than HuGE Navigator. A literature search with PubMed did not reveal any additional studies. The GWAS covered a range of responses, namely: adverse motor side effects (EPS), metabolic changes [including increases in cholesterol, triglycerides and body mass index (BMI)], changes in neurocognitive functioning, changes in symptom severity rated both by the patient and clinician, adverse cardiac symptoms, and treatment-refractoriness. The identified GWAS

found a combined total of 62 SNPs significantly associated with antipsychotic treatment response, as shown in Table 3.3. These variants are ranked by earliest to most recent study and, within this ranking, by most significant *P*-value. Considering the location of the 62 variants identified, the most commonly occurring region is 4q24, with five SNPs at this locus having reached significance in four different GWAS. One of these variants, rs230529, was significant for treatment-refractoriness (Liou *et al.*, 2012), whilst the other four – rs7669317, rs1405687, rs2636697 and rs2636719 – were associated with EPS, metabolic side effects, and clinician-rated symptom severity respectively (Åberg *et al.*, 2010; Adkins *et al.*, 2011; Clark *et al.*, 2013). Most of the GWAS indicated the SNP position in terms of the closest gene, without specifying whether it is located in an intron or whether it is intergenic. Once annotated with SeattleSeq Annotation 137, it was shown that only one SNP, rs17727261, occurs in a protein-coding region, namely in an exon of the contactin associated protein-like 5 (*CNTNAP5*) gene. Of the other SNPs, 27 are intronic, one occurs in the 3'-UTR of the zinc finger protein 202 (*ZNF202*) gene, and the remaining 33 are intergenic. As shown in bold in Table 3.3, there were four instances in which SeattleSeq classified SNPs as intergenic whilst the GWAS listed them as occurring within a gene (Clark *et al.*, 2013).

3.4.2. GWAS cohort ancestry and LD analyses

The 62 significant SNPs were analysed for variants in linkage disequilibrium according to the 1000 Genomes and HapMap population groups. Populations were selected according to the ancestral make-up of each GWAS, as indicated in Table 3.4. All GWAS were corrected for ancestry, in cases where the cohort was comprised of more than one ethnic group. Additionally, many of the studies performed subsample testing in order to ascertain whether a particular population group was driving a significant association. Except for a few instances in which a SNP was invariant in a particular subgroup, the majority of SNPs were significant – albeit to varying degrees – across population subgroups. Therefore, to be as inclusive as possible, no population groups were excluded for further LD analysis.

There were several instances in which a variant found on SNAP was in LD ($r^2 \geq 0.8$) with more than one GWAS SNP. No SNPs from independent GWAS studies shared LD for any population groups, although there were significant SNPs in LD *within* specific GWAS. For example, rs2636697, and rs2636719 from the Clark *et al.* study (2013) are in perfect LD ($r^2 = 1$ for 1000 Genomes CEU), as well as rs7105881, rs7119817, and rs7108821 ($r^2 = 1$ for each pair for 1000 Genomes CEU) from the Adkins *et al.* GWAS (2011). Once all variants for all relevant populations were combined, as described in 3.3.2, there was a total of 535 unique SNPs, including the original 62 GWAS SNPs.

Table 3.3: Significant SNPs from antipsychotic pharmacogenomic GWAS identified by HuGE Navigator and the NHGRI GWAS Catalog.

STUDY	SAMPLE SIZE Initial, Replication	VARIANT	LOCUS	GENE ^a		RESPONSE MEASUREMENT	P-VALUE	EFFECT ^b
				Designated	Mapped			
Åberg <i>et al.</i> , 2010	738 ^c	rs17022444	2p12	None	Intergenic	EPS (SAS)	1 x 10 ⁻¹⁰	+
		rs7669317	4q24	None	Intergenic	EPS (AIMS)	8 x 10 ⁻⁸	+
		rs2126709	11q24.1	<i>ZNF202</i>	<i>ZNF202</i> (3'-UTR)	EPS (SAS)	4 x 10 ⁻⁷	+
Adkins <i>et al.</i> , 2011	738 ^c	rs1568679	15q14	<i>MEIS2</i>	<i>MEIS2</i> (intron)	Hip circumference	1 x 10 ⁻⁸	+
		rs1967256	5q14.3	<i>GPR98</i>	<i>GPR98</i> (intron)	Haemoglobin A1c	3 x 10 ⁻⁸	+
		rs11954387	5q14.3	<i>GPR98</i>	<i>GPR98</i> (intron)	Haemoglobin A1c	3 x 10 ⁻⁸	+
		rs1405687	4q24	None	Intergenic	Hip circumference	5 x 10 ⁻⁸	-
		rs1568679	15q14	<i>MEIS2</i>	<i>MEIS2</i> (intron)	Waist circumference	6 x 10 ⁻⁸	+
		rs13224682	7p22.3	<i>PRKAR2B</i>	<i>PRKAR2B</i> (intron)	Triglycerides	6 x 10 ⁻⁸	+
		rs1464500	12p12.1	<i>SOX5</i>	<i>SOX5</i> (intron)	HDL cholesterol	1 x 10 ⁻⁷	+
		rs17651157	18q12.2	<i>FHOD3</i>	<i>FHOD3</i> (intron)	Triglycerides	1 x 10 ⁻⁷	+
		rs6735179	2p25.3	None	Intergenic	Triglycerides	1 x 10 ⁻⁷	+
		rs518590	13q12.11	None	Intergenic	HDL cholesterol	2 x 10 ⁻⁷	+
		rs10502661	18q12.2	<i>FHOD3</i>	<i>FHOD3</i> (intron)	Triglycerides	2 x 10 ⁻⁷	+
		rs1187614	14q32.13	<i>CLMN</i>	<i>CLMN</i> (intron)	Total cholesterol	2 x 10 ⁻⁷	-
		rs6741819	2p25.1	<i>RNF144A</i>	<i>RNF144A</i> (intron)	Triglycerides	2 x 10 ⁻⁷	+
		rs4838255	9q33.1	<i>ASTN2</i>	<i>ASTN2</i> (intron)	Triglycerides	3 x 10 ⁻⁷	+
		rs2994684	10p11.22	None	Intergenic	Triglycerides	3 x 10 ⁻⁷	+
		rs977396	8q22.3	None	Intergenic	Total cholesterol	3 x 10 ⁻⁷	+
		rs7105881	11q23.1	None	Intergenic	Hip circumference	3 x 10 ⁻⁷	+
		rs1117324	2p24.1	None	Intergenic	Hip circumference	3 x 10 ⁻⁷	+
		rs4783227	16q23.3	None	Intergenic	Total cholesterol	4 x 10 ⁻⁷	-
		rs320209	9q31.1	None	Intergenic	Glucose	4 x 10 ⁻⁷	+
rs7108821	11q23.1	None	Intergenic	Hip circumference	4 x 10 ⁻⁷	+		
rs10499504	7p21.1	None	Intergenic	Total cholesterol	4 x 10 ⁻⁷	-		
rs7119817	11q23.1	None	Intergenic	Hip circumference	5 x 10 ⁻⁷	+		
rs9658108	6p21.31	<i>PPARD</i>	<i>PPARD</i> (intron)	Glucose	5 x 10 ⁻⁷	+		
rs17100498	5q31.3	None	Intergenic	Haemoglobin A1c	5 x 10 ⁻⁷	+		

STUDY	SAMPLE SIZE Initial, Replication	VARIANT	LOCUS	GENE ^a		RESPONSE MEASUREMENT	P-VALUE	EFFECT ^b
				Designated	Mapped			
Adkins <i>et al.</i> , 2011	738 ^c	rs399885	2p12	None	Intergenic	Heart rate	5 x 10 ⁻⁷	+
McClay <i>et al.</i> , 2011a	738 ^c	rs286913	11p13	<i>EHF</i>	<i>EHF</i> (intron)	Neurocognition: vigilance	7 x 10 ⁻⁸	-
		rs11240594	1q32.1	<i>SLC26A9</i>	<i>SLC26A9</i> (intron)	Neurocognition: processing speed	1 x 10 ⁻⁷	-
		rs11110077	12q23.1	<i>ANKS1B</i>	<i>ANKS1B</i> (intron)	Neurocognition: working memory	4 x 10 ⁻⁷	+
		rs7520258	1q42.3	<i>GPR137B</i>	<i>GPR137B</i> (intron)	Neurocognition: working memory	5 x 10 ⁻⁷	+
		rs12726652	1p13.3	None	Intergenic	Neurocognition: working memory	5 x 10 ⁻⁷	+
		rs11214606	11q23.2	<i>DRD2</i>	<i>DRD2</i> (intron)	Neurocognition: working memory	5 x 10 ⁻⁷	+
McClay <i>et al.</i> , 2011b	738 ^c	rs2833556	21q22.11	<i>HUNK</i>	<i>HUNK</i> (intron)	Neurocognition: reasoning	5 x 10 ⁻⁷	-
		rs17390445	4p15.1	None	Intergenic	Positive symptoms (PANSS)	1 x 10 ⁻⁷	+
		rs888219	9q33.3	None	Intergenic	Negative symptoms (PANSS)	2 x 10 ⁻⁷	-
		rs7968608	12q23.1	<i>ANKS1B</i>	<i>ANKS1B</i> (intron)	Negative symptoms (PANSS)	3 x 10 ⁻⁷	-
		rs17727261	2q14.3	<i>CNTNAP5</i>	<i>CNTNAP5</i> (exon)	Negative symptoms (PANSS)	5 x 10 ⁻⁷	-
Åberg <i>et al.</i> , 2012	738 ^c	rs11722719	4p15.1	None	Intergenic	Positive symptoms (PANSS)	5 x 10 ⁻⁷	+
		rs4959235	6p25.2	<i>SLC22A23</i>	<i>SLC22A23</i> (intron)	QTc interval prolongation	2 x 10 ⁻⁷	+
Athanasias <i>et al.</i> , 2012	594	rs10458561	1p31.1	None	Intergenic	QTc interval prolongation	4 x 10 ⁻⁷	+
		rs7838490	8q21.3	None	Intergenic	BMI	6 x 10 ⁻⁸	+
Liou <i>et al.</i> , 2012	522 cases and 806 controls, 273 cases	rs11615274	12q21.1	None	Intergenic	HDL cholesterol	9 x 10 ⁻⁸	-
		rs230529	4q24	<i>NFKB1</i>	<i>NFKB1</i> (intron)	Treatment-refractory schizophrenia	2 x 10 ^{-7 d}	+
		rs11265461	1q23.3	None	Intergenic		2 x 10 ^{-7 d}	+
rs10218843	1q23.3	None	Intergenic	3 x 10 ^{-7 d}	+			
Malhotra <i>et al.</i> , 2012	139 73, 40, 92	rs489693	18q21.32	None	Intergenic	Severe weight gain, several other metabolic indices	6 x 10 ^{-12 d}	+
Clark <i>et al.</i> , 2013	738 ^c	rs8050896	16q22.1	None	Intergenic	Clinical global impression severity scale (CGI-S)	4 x 10 ⁻⁸	-
		rs17382202	5q12.1	<i>PDE4D</i>	<i>PDE4D</i> (intron)	Patient global impression (PGI) scale	4 x 10 ⁻⁸	-
		rs10170310	2q22.1	<i>SPOPL</i>	<i>SPOPL</i> (intron)	PGI	1 x 10 ⁻⁷	+
		rs6688363	1q23.2	<i>ATP1A2</i>	Intergenic	CGI-S	2 x 10 ⁻⁷	+
		rs7395555	11q14.1	None	Intergenic	CGI-S	2 x 10 ⁻⁷	-

STUDY	SAMPLE SIZE Initial, Replication	VARIANT	LOCUS	GENE ^a		RESPONSE MEASUREMENT	P-VALUE	EFFECT ^b
				Designated	Mapped			
Clark <i>et al.</i> , 2013	738 ^c	rs17742120	5q12.1	<i>PDE4D</i>	<i>PDE4D</i> (intron)	PGI	2 x 10 ⁻⁷	-
		rs2164660	5q12.1	<i>PDE4D</i>	<i>PDE4D</i> (intron)	PGI	2 x 10 ⁻⁷	-
		rs711355	15q13.1	<i>TJP1</i>	Intergenic	PGI	2 x 10 ⁻⁷	-
		rs2980976	18q21.3	<i>TNFRSF11A</i>	Intergenic	CGI-S	3 x 10 ⁻⁷	+
		rs2636697	4q24	<i>PPA2</i>	<i>PPA2</i> (intron)	CGI-S	4 x 10 ⁻⁷	+
		rs2636719	4q24	<i>PPA2</i>	<i>PPA2</i> (intron)	CGI-S	5 x 10 ⁻⁷	+
		rs785423	15q13.1	<i>TJP1</i>	Intergenic	PGI	5 x 10 ⁻⁷	-
		rs813676	15q13.1	<i>TJP1</i>	Intergenic	PGI	5 x 10 ⁻⁷	-

^a Gene designated to the SNP by GWAS authors vs. gene mapped by SeattleSeq Annotation 137 (differences in bold).

^b Direction of effect of minor allele, where “+” denotes minor allele frequency (MAF) associated with poorer response or presence of ADR.

^c Identical cohort from the CATIE study (Lieberman *et al.*, 2005).

^d Joint probability from meta-analysis of initial and replication cohorts.

Shaded values indicate SNPs with $P \leq 5 \times 10^{-8}$.

SAS = Simpson-Angus Scale; AIMS = Abnormal Involuntary Movement Scale; UTR = untranslated region; Haemoglobin A1c = glycohaemoglobin (used to measure plasma glucose levels); HDL = high-density lipoprotein; QTc = interval between ventricular depolarisation (Q wave) and repolarisation (T wave) in electrocardiogram, corrected for heart rate.

Table 3.4: Ancestry breakdown of the four cohorts studied by relevant GWAS and corresponding SNAP populations included for LD analysis (<http://www.broadinstitute.org/mpg/snap/>).

Study	Sample ancestry proportions		Included SNAP populations
	Initial	Replication	
Åberg <i>et al.</i> , 2010 Adkins <i>et al.</i> , 2011 McClay <i>et al.</i> , 2011a McClay <i>et al.</i> , 2011b Åberg <i>et al.</i> , 2012 Clark <i>et al.</i> , 2013	57% EA, 29% AA, 14% other	None	All
Athanasu <i>et al.</i> , 2012	100% Caucasian	None	CEU, TSI, CEU+TSI
Liou <i>et al.</i> , 2012	100% Han Chinese	100% Han Chinese	CHBJPT, CHD, JPT+CHB+CHD
Malhotra <i>et al.</i> , 2012	55% Caucasian, 23% AA, 22% other	70% Caucasian, 30% AA; and 100% Caucasian	All

EA = European American; AA = African American; CEU = Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain (CEPH) collection; TSI = Tuscans in Italy; CHBJPT = Han Chinese in Beijing, China, and Japanese in Tokyo, Japan; CHD = Chinese in metropolitan Denver, Colorado; JPT = Japanese in Tokyo, Japan; CHB = Han Chinese in Beijing, China.

3.4.3. RegulomeDB analysis

The total of 535 variants was analysed with RegulomeDB and each SNP was assigned a score based on its regulatory potential. Twenty SNPs scored significantly (≤ 3), and are listed in Table 3.5 together with their predicted regulatory effect. The results include two original GWAS SNPs, namely rs6741819 (Adkins *et al.*, 2011) and rs10458561 (Åberg *et al.*, 2012). The remaining 18 SNPs were identified from LD analyses. The top 10 results obtained a score of one, meaning that they have shown to act as eQTLs, altering the expression of the following genes: mannosidase, beta A, lysosomal (*MANBA*), collagen, type IX, alpha 2 (*COL9A2*), and DEAD/H box helicase 11 (*DDX11*). Nine of these SNPs were in LD with rs230529 from the Liou *et al.* (2012) study, and rs10492354 was in LD with a significant SNP identified by McClay *et al.* (2011b). Proteins, such as TFs, that bind in the region of the SNP are indicated in Table 3.5. Additionally, their corresponding motifs and predicted position weight matrices (PWMs) are shown. All 20 SNPs show evidence for changes in chromatin state and histone modifications at their particular locus. A full list of these can be obtained for each SNP by entering the rs number into the database (<http://regulome.stanford.edu/>).

Since 14 of the 20 SNPs occur in the 4q24 region, this locus was further investigated with the use of the University of California, Santa Cruz (UCSC) ENCODE browser (<http://genome.ucsc.edu/ENCODE/>). As shown in Figure 3.1, there are several lines of

evidence across different cell types that point to regulatory function in this region, including TFBS, histone marks and open chromatin, DNase I hypersensitive sites, and the start of transcription of the nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (*NFKB1*) gene.

3.4.4. rSNPBase analysis



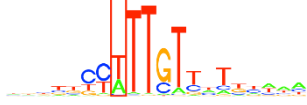



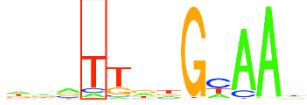

Variants were also analysed with rSNPBase. This tool predicted that 222 of the 535 SNPs affected regulation, either proximally, distally, or post-transcriptionally. Additionally, it identified the predicted affected genes. Table 3.6 shows the 16 SNPs that were deemed significant by both rSNPBase and RegulomeDB. Motifs, genes, proteins, affected genes, and eQTLs common to both sets of results are highlighted in the table.

3.4.5. Variants affecting binding motifs

Changes in motif binding affinity were assessed with tRap and the most significant results are shown in Table 3.7. A total of 111 of the 535 variants were predicted by tRap to alter binding affinity, with many SNPs affecting more than one motif. This table includes the ten SNPs that are predicted to cause the largest decreases and increases in affinity respectively, measured by the log ratio. Overlaps with rSNPBase or RegulomeDB results are highlighted. The variant rs10492354 was predicted to be an eQTL for *DDX11* expression by RegulomeDB, and caused the second highest decrease in binding affinity, at the breast cancer type I susceptibility protein (BRCA1) motif. Secondly, rs230493 was predicted by tRap to increase binding significantly at the GATA binding protein 2 (GATA2) motif, as well as show eQTL evidence and distal and post-transcriptional regulatory effects by RegulomeDB and rSNPBase. The GATA2 motif is the only motif that overlaps with significant bound proteins or motifs predicted by RegulomeDB or rSNPBase.

The variants implicated in the control traits (response to antidepressants, response to hepatitis C treatment, and eye colour) were also analysed with tRap. Extensive overlap of motifs between antipsychotic response SNPs and the control SNPs was found. Table S1 shows that only 12 motifs were unique to antipsychotic response, with a maximum of two SNPs significantly predicted to affect binding of each motif.

Table 3.5: Top predicted rSNPs from RegulomeDB with associated regulatory targets and effects (<http://regulome.stanford.edu/>).

SNP	Position (GRCh37/ hg19)	Score ^a	eQTL target(s)	Bound protein(s)	Sequence motif(s)		Other evidence	
					Name	Position weight matrix (PWM) ^b	Chromatin changes	Histone modifications
rs3774959	4:103511113	1b	MANBA, COL9A2	RFX3	Lmo2complex		Yes	Yes
rs230505	4:103481350	1d	MANBA, COL9A2	BATF	E12, Six4, Myf6, MyoD		Yes	Yes
rs230532	4:103450166	1f	MANBA, COL9A2	None	Nanog		Yes	Yes
rs230520	4:103465611	1f	MANBA, COL9A2	None	None		Yes	Yes
rs1599961	4:103443568	1f	MANBA, COL9A2	None	None		Yes	Yes
rs230504	4:103481560	1f	MANBA, COL9A2	None	None		Yes	Yes
rs230493	4:103486215	1f	MANBA, COL9A2	RFX3	Six-1		Yes	Yes
rs747559	4:103414174	1f	MANBA, COL9A2	None	Pitx2, Cdc5		Yes	Yes
rs4648055	4:103515312	1f	MANBA, COL9A2	FOS	None		Yes	Yes
rs10492354	12:31357101	1f	DDX11	None	FOXP1, Foxk1		Yes	Yes
rs3774933	4:103426338	2a	None	FOXA1, CEBPB	C/EBP		Yes	Yes
rs6741819 ^c	2:7147972	2b	None	EBF1, GATA1, CTCF	CNOT3		Yes	Yes

SNP	Position (GRCh37/ hg19)	Score ^a	eQTL target(s)	Bound protein(s)	Sequence motif(s)		Other evidence	
					Name	Position weight matrix (PWM) ^b	Chromatin changes	Histone modifications
rs230526	4:103458824	2b	None	AF	Barhl-1 , Isl2, Arid3a, Barhl2, Dbx2, Lhx1, Lhx3, Lhx5, Lmx1a, Lmx1b, Msx1, Sox17, Sox8, Tlx2		Yes	Yes
rs17032850	4:103507702	2b	None	SPI1, EBF1, BATF	TCF-4 , HNF1		Yes	Yes
rs2272676	4:103423325	2b	None	POLR2A, TCF4, ZNF263, CDX2, SPI1, TBP, JUNB, NFKB1	BLIMP1 , EWSR1-FLI1		Yes	Yes
rs1352318	8:89566902	2c	None	CTCF, RAD21	CTCF		Yes	Yes
rs10458561 ^c	1:70921172	3a	None	EP300	C/EBPalpha		Yes	Yes
rs230495	4:103487299	3a	None	MAFK	AP-1 , Nanog		Yes	Yes
rs17440909	5:58957830	3a	None	HNF4G, SP1, EP300, TCF4	AML		Yes	Yes
rs66919541	12:73660166	3a	None	GATA2	Zfp740		Yes	Yes

^a Score definitions are listed in Figure 3.1.

^b PWM corresponds to motif in bold. Red box indicates SNP position.

^c SNP from original GWAS study.

Full list of other evidence such as histone marks and open chromatin for specific cell types can be obtained by entering the rs number into RegulomeDB.

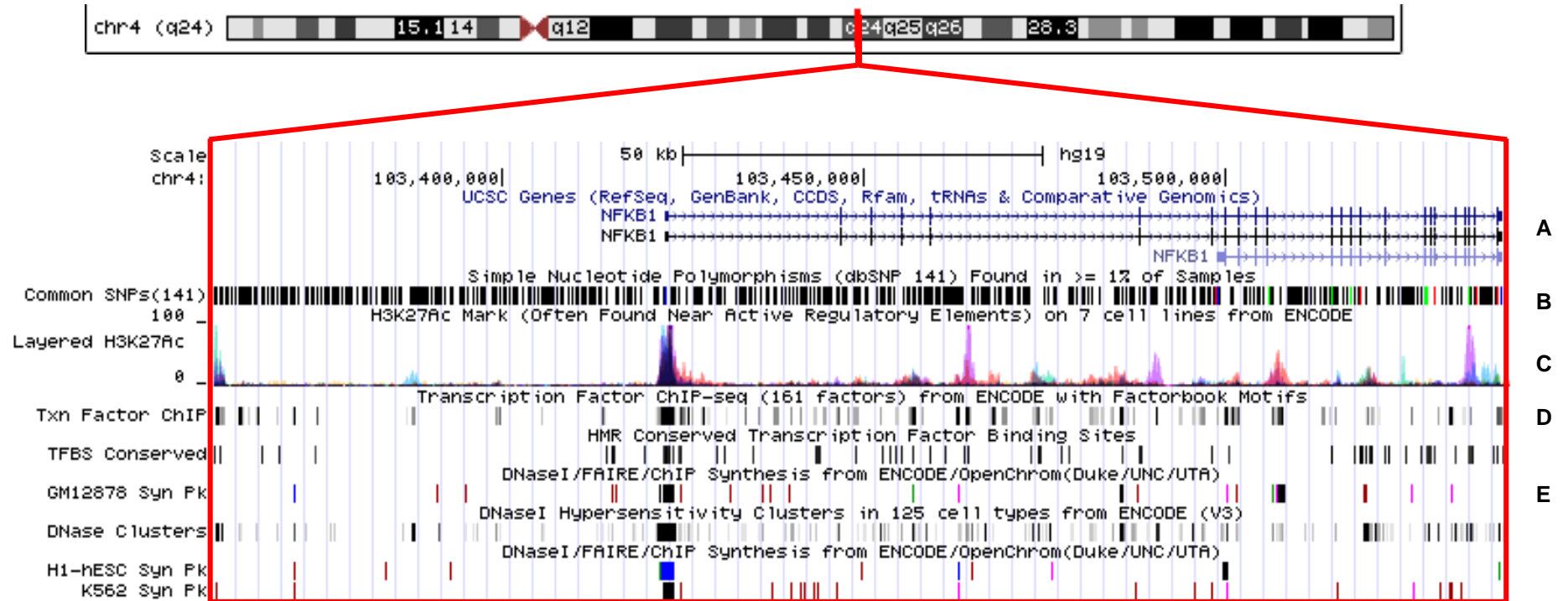


Figure 3.1: Magnified view of the 4q24 genomic region on the UCSC Genome Browser with ENCODE data tracks (<http://genome.ucsc.edu/ENCODE/>). This region contains 14 of 20 predicted rSNPs according to RegulomeDB. ENCODE tracks show **A**) transcription start of *NFKB1* gene, **B**) common SNPs (MAF > 1%) identified in this region, **C**) peaks for histone mark H3K27Ac, associated with open chromatin, **D**) TFBS determined by ChIP-Seq and **E**) DNaseI hypersensitive sites determined by various experiments. Different colours indicate evidence in different cell lines.

Table 3.6: rSNPBase annotations for significant RegulomeDB SNPs, arranged by genomic position (<http://rsnp.psych.ac.cn/>).

SNP	eQTL target(s)	Type of regulation			Affected gene(s)
		Proximal: TF	Post-transcriptional: RNA-binding protein	Distal	
rs6741819	SOCS3, RNF144A		ELAVL1, IGF2BP1, PABPC1, CELF1	Yes	RNF144A, RNF144-AS1
rs747559	COL9A2, MANBA, MAPKSP1			Yes	NFKB1
rs2272676		Max, Rad21, BCLAF1, ELF1, FOXM1, Pol2-4H8, Pol2, NFKB, ZNF263, CTCF, EBF1, several others*	PABPC1	Yes	NFKB1, FAM173B, CCT5, BCL7B, UBE2D3, SLC39A8, CISD2, SLC9B1, WDR74
rs3774933		CHD1, MafK, ARID3A, CEBPB	PABPC1		NFKB1
rs1599961	COL9A2, MANBA, MAPKSP1		PABPC1	Yes	NFKB1
rs230532	CCNG2, COL9A2, MANBA, ADH7, BANK1, CISD2		PABPC1	Yes	NFKB1
rs230526	MAPKSP1, C18orf21		PABPC1		NFKB1
rs230520	PLEKHA4, COL9A2, MANBA, ADH7, BANK1, CISD2		PABPC1	Yes	NFKB1
rs230505	TMED2, MANBA, COL9A2		PABPC1		NFKB1
rs230504	ING2, COL9A2, MANBA, ADH7, CISD2		PABPC1		NFKB1
rs230493	CACNB1, COL9A2, MANBA		PABPC1	Yes	NFKB1
rs230495		MafF, MafK	PABPC1	Yes	NFKB1
rs17032850			PABPC1	Yes	NFKB1
rs3774959	WDR27, MANBA, COL9A2	Pol2-4H8	PABPC1		NFKB1
rs4648055	CKAP4, COL9A2, MANBA, ADH7, CISD2		PABPC1		NFKB1
rs17440909			ELAVL1		PDE4D

Shaded genes, proteins and motifs indicate commonalities between rSNPBase and RegulomeDB SNPs.

* Many other TFs are listed on rSNPBase, but all TFs in common with RegulomeDB for rs2272676 are listed in the table.

Table 3.7: Top 10 SNPs predicted by TRAP to increase or decrease motif binding affinity significantly.

Decreases in binding affinity				
SNP	log ratio	Jaspar motif ID	Motif	Total SNPs^a
rs2367184	-3.78665	MA0133.1	BRCA1	9
rs10492354	-3.60905	MA0133.1	BRCA1	9
rs1585215	-3.27794	MA0084.1	SRY	8
rs60371688	-2.70746	MA0133.1	BRCA1	9
rs11613776	-2.48881	MA0036.1	GATA2	4
rs246430	-2.47982	MA0042.1	FOXI1	1
rs67609022	-2.09686	MA0098.1	ETS1	9
rs11214606	-2.09355	MA0098.1	ETS1	9
rs3113628	-2.0879	MA0098.1	ETS1	9
rs56293675	-2.08772	MA0098.1	ETS1	9
Increases in binding affinity				
SNP	log ratio	Jaspar motif ID	Motif	Total SNPs^a
rs1020760	3.872955	MA0069.1	Pax6	6
rs4648052	2.659042	MA0084.1	SRY	8
rs71526953	2.337937	MA0133.1	BRCA1	9
rs17742544	2.240923	MA0043.1	HLF	6
rs230493	1.968377	MA0036.1	GATA2	4
rs58133638	1.777604	MA0095.1	YY1	2
rs62328542	1.725868	MA0099.2	JUN::FOS ^b	2
rs55999909	1.715268	MA0259.1	HIF1A::ARNT ^b	1
rs111577254	1.700943	MA0098.1	ETS1	9
rs821102	1.656439	MA0258.1	ESR2	1

^a Total number of SNPs predicted to increase or decrease binding affinity significantly for each motif in the study.

^b Heterodimer

Shaded SNPs and motifs were also deemed significant by RegulomeDB and/or rSNPBase.

3.4.6. Nonsynonymous coding SNPs

PolyPhen-2 was utilised to assess functional implications of the 535 SNPs with regards to protein-coding regions. One SNP located in an exon of the *CNTNAP5* gene, rs17727261, was classified as nonsynonymous. This missense variant, identified as significant in the McClay *et al.* (2011b) GWAS, results in a serine to leucine substitution at codon 452 (S452L). Polyphen-2 predicted this variant to be benign, with a score of 0.011. Therefore, none of the SNPs are expected to have damaging effects on protein products. This is expected given that the majority are in noncoding regions.

3.4.7. Affected genes and pathways

A list of affected genes was compiled by combining eQTL targets from rSNPBase and RegulomeDB, and affected genes predicted by rSNPBase. *CNTNAP5* was also included, for

a total of 126 genes. After excluding pseudogenes and unprocessed transcripts, the list consisted of 118 unique genes, listed in Table S2. This gene set was uploaded to DAVID for functional annotation. Four pathways were identified by DAVID and are ranked by *P*-value in Table 3.8. Terms with larger fold change values (> 1.5) and smaller uncorrected *P*-values (< 0.1) should be considered significant for further investigation (Huang *et al.*, 2009). Therefore, all four pathways identified are of interest. *NFKB1*, affected by a total of 104 rSNPs, plays a role in three of these pathways, namely HIV-I negative factor (Nef) functioning, chronic myeloid leukaemia, and human cytomegalovirus (CMV) and mitogen-activated protein kinase (MAPK) pathways. The retinoblastoma 1 (*RB1*) gene also occurs in all three of these pathways. The fourth pathway, platelet-derived growth factor (PDGF) signalling, contains four genes affected by a total of six rSNPs.

Table 3.8: Pathways identified by DAVID for affected genes (<http://david.abcc.ncifcrf.gov/home.jsp>).

Tool	Pathway	Fold change	<i>P</i> -value	Genes	Total SNPs*
BIOCARTA	HIV-I Nef	7.4	0.05	<i>ACTG1</i>	1
				<i>NFKB1</i>	104
				<i>RB1</i>	1
PANTHER	PDGF signalling pathway	3.7	0.08	<i>RAB25</i>	2
				<i>SHC1</i>	2
				<i>EHF</i>	1
				<i>PKN2</i>	1
KEGG	Chronic myeloid leukaemia	5.8	0.09	<i>SHC1</i>	2
				<i>NFKB1</i>	104
BIOCARTA	Human CMV and MAPK pathways	18.7	0.09	<i>RB1</i>	1
				<i>NFKB1</i>	104

* Total SNPs predicted by RegulomeDB and/ or rSNPBase to affect particular gene.

Nef = negative factor; PDGF = platelet-derived growth factor; CMV = cytomegalovirus; MAPK = mitogen-activated protein kinase.

The genes for each pathway were uploaded to GeneMANIA for further analysis. Networks, as well as corrected probability scores for common functions between genes, were assessed. Figure 3.2 shows the network for genes involved in the HIV-I Nef pathway. Two of the three genes in this pathway, *NFKB1* and the actin, gamma 1 (*ACTG1*) gene, were identified by GeneMANIA to play a role in immune response-activating cell surface receptor signalling and Fc receptor signalling, with false discovery rate (FDR) values of 0.07 and 0.08 respectively.

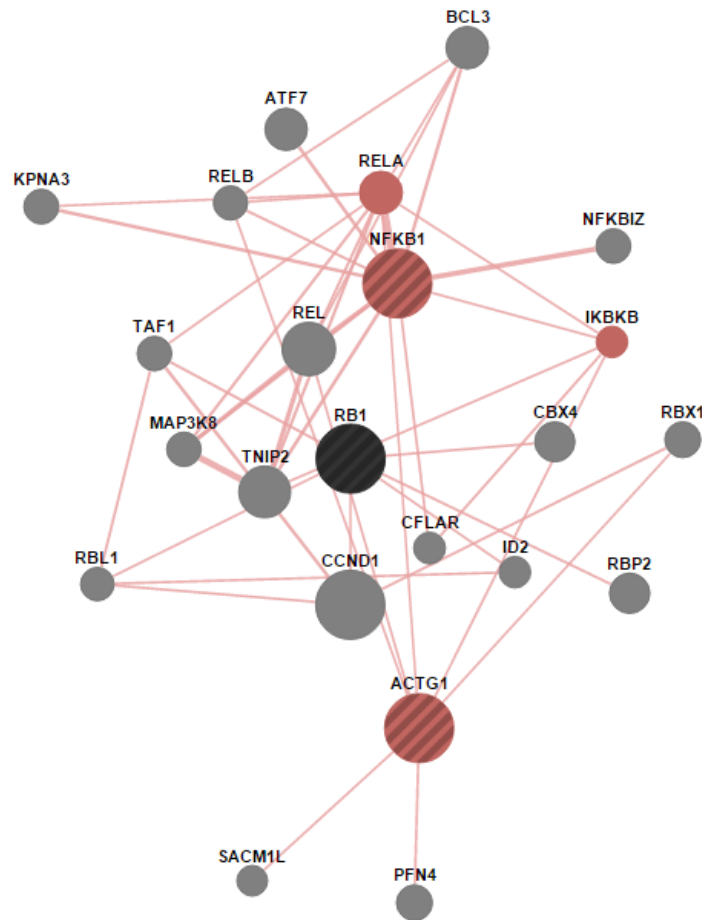


Figure 3.2: GeneMANIA network for affected genes in the HIV-1 Nef pathway according to DAVID (*ACTG1*, *NFKB1* and *RB1*), with related genes in grey. Red genes are involved in immune response-activating cell surface receptor signalling and Fc receptor signalling pathways. Red connections indicate physical interactions (<http://www.genemania.org/>).

Additionally, Figure 3.3 illustrates the affected genes involved in the chronic myeloid leukaemia pathway. No functions common to all three genes in this pathway were found, however there were several functional overlaps between *NFKB1* and the Src homology 2 domain containing (SHC) transforming protein 1 (*SHC1*) gene. These include Fc receptor signalling (FDR = 0.001) and neurotrophin signalling (FDR = 0.002), as well as the previously mentioned functions that are common to *NFKB1* and *ACTG1*. Connections between genes involved in the Human CMV and MAPK pathways (*NFKB1* and *RB1*) can also be viewed in Figure 3.3 (B). These genes both take part in regulatory region DNA binding (FDR = 0.015) as well as the regulation of lipid metabolism (FDR = 0.030). The network for PDGF signalling did not show functions common to any of the four genes and thus is not shown.

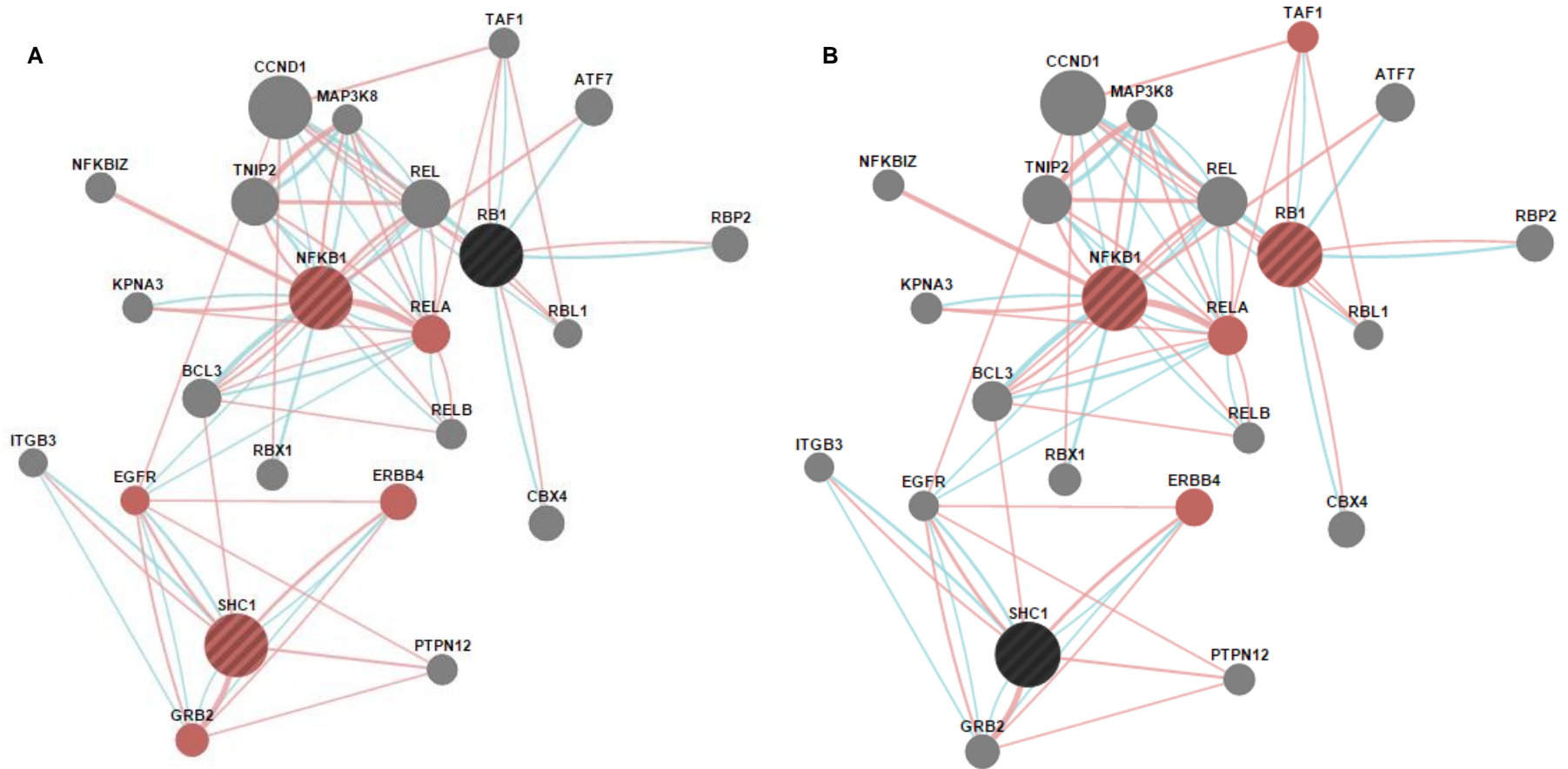


Figure 3.3: GeneMANIA networks indicating genes in the chronic myeloid leukaemia pathway, and human CMV and MAPK pathways according to DAVID, with related genes in grey. Red genes in **A)** are involved in Fc receptor signalling and neurotrophin signalling pathways, and red genes in **B)** play a role in DNA binding in regulatory regions. Red connections indicate physical interactions and blue connections indicate pathways (<http://www.genemania.org/>).

3.4.8. Tissue-specific gene expression

Expression patterns of the ten genes affected by the most number of SNPs were analysed with FANTOM5. The highest brain- and liver-related expression values, as well as highest overall expression value, were measured as shown in Table 3.9. For four of the affected genes, indicated in bold, the highest overall expression was in a cell line related to the brain, i.e. the pineal gland, globus pallidus, and astrocytes in the cerebellum. A few genes showed negligible expression (TPM < 10) in either brain or liver cells, with the alcohol dehydrogenase class 4 mu/sigma chain (*ADH7*) gene showing no expression in either. Of the ten genes, the highest liver-related expression was demonstrated by calmin (calponin-like, transmembrane) (*CLMN*) in hepatocytes (TPM = 81.6).

Table 3.9: Brain- and liver-specific expression of ten most affected genes according to FANTOM5 (<http://fantom.gsc.riken.jp/5/sstar>).

Gene	Total SNPs	Highest brain-related expression		Highest liver-related expression		Highest overall expression	
		Tissue	TPM ^a	Tissue	TPM ^a	Tissue	TPM ^a
<i>NFKB1</i>	104	Dura mater	32.1	Hepatic sinusoidal endothelial cells	33.2	CD14+ monocytes	1314.2
<i>PDE4D</i>	25	Parietal lobe	27.9	Foetal liver	6.6	CD19+ B cells	68.3
<i>MANBA</i>	23	Brain smooth muscle	44.2	Hepatic mesenchymal stem cells	35.2	Neutrophils	290.5
<i>GPR98</i>	18	Pineal gland	138.2	None	0.0	Pineal gland	138.2
<i>COL9A2</i>	18	Cerebellum astrocytes	191.2	Hepatic stellate cells (lipocytes)	5.9	Cerebellum astrocytes	191.2
<i>CLMN</i>	14	Globus pallidus	138.6	Hepatocytes	81.6	Globus pallidus	138.6
<i>DICER1</i>	10	Globus pallidus	53.5	Hepatocytes	36.8	Globus pallidus	53.5
<i>CISD2</i>	9	Meningeal cells	39.7	Adult liver	70.5	Reticulocytes	193.6
<i>SPOPL</i>	6	Globus pallidus	21.0	Foetal liver	10.6	Neutrophils	65.8
<i>ADH7</i>	6	None	0.0	None	0.0	Oesophagus	130.6

^a TPM = tags per million; TPM ≥ 10 represents active gene expression.

Cases in which relevant tissue expression is also highest overall expression are indicated in bold.

3.5. Discussion

This study aimed to assess the functional impact of SNPs implicated in previous antipsychotic response GWAS, in order to improve our understanding of the mechanisms behind antipsychotics and the genetics of treatment response. The novel and comprehensive bioinformatics pipeline provides an approach that includes the analysis of

predicted and experimentally validated regulatory regions overlapping the SNPs, and assesses the regulatory potential of each variant. Furthermore, functional interpretations can be made by exploring the predicted regulated genes, implicated pathways, and expression in relevant tissue types. These factors allow for hypothesis-free evaluation of significant genomic variants, and thus could uncover novel genes and pathways associated with treatment response.

3.5.1. Antipsychotic response GWAS

3.5.1.1. GWAS study design

Firstly, the shortage of studies on antipsychotic treatment response is apparent. For example, when the interactive version of the NHGRI GWAS Catalog is consulted (<http://www.ebi.ac.uk/fgpt/gwas/>), one can see the plethora of significant SNPs for other traits or disorders, such as those of the immune or digestive systems. Using the Catalog's genome-wide significance cut-off of $P \leq 5 \times 10^{-8}$, only seven SNPs in Table 3.3 qualify as significant. Even with a more lenient threshold of $P \leq 5 \times 10^{-7}$, only 62 SNPs from a total of nine studies are deemed important. However, this is not surprising since, to date, only twelve GWAS on antipsychotic response have been performed, and they do not match the large-scale studies like those conducted by the PGC (Ripke *et al.*, 2013; 2014). Additionally, the generally weak association signals contrast with those found in schizophrenia susceptibility GWAS, for which the strongest association signal has a P -value of 1.47×10^{-16} (Strange *et al.*, 2012). The reason for relatively weak associations can be explained by the limited sample size of previous treatment response GWAS, as well as the difficulty in obtaining a well-characterised and clinically homogeneous cohort.

The CATIE cohort, consisting of 738 genotyped individuals, was tested for associations with different response traits for six of the nine studies, as shown in Table 3.3. This cohort is both the largest and most thoroughly assessed antipsychotic response trial that has patient DNA available (Adkins *et al.*, 2011). Individuals in CATIE began treatment on either one of four SGAs (ziprasidone, risperidone, quetiapine or olanzapine), or the FGA perphenazine. If treatment was deemed ineffective, patients could either go on to clozapine or a different SGA. The main aim of this multiphase, randomised trial was to compare effectiveness of different SGAs, and measure the efficacy of SGAs against perphenazine (Stroup *et al.*, 2003). The comprehensive clinical data obtained in the CATIE trial is useful for assessing different aspects of treatment response. The range of outcomes assessed include adverse metabolic and cardiovascular measures, EPS, rated with three different scales, general

cognitive and psychosocial functioning, and treatment response measured with the PANSS, PGI and CGI-S (Lieberman *et al.*, 2005).

Although differences in treatment efficacy are still under debate (Lewis and Lieberman, 2008; Meltzer, 2013), these agents have demonstrated significantly different side effect profiles, particularly in the case of clozapine (Brandl *et al.*, 2014). Our understanding of the mechanisms of antipsychotics is limited, but it is likely that different adverse reactions indicate different biological pathways, and thus antipsychotic heterogeneity limits statistical power in GWAS (Ni *et al.*, 2013). In order to improve this, five of the CATIE GWAS increased clinical homogeneity at the price of decreasing sample size; in other words, the authors performed drug-specific GWAS on patient subgroups. For example, the two significant SNPs identified by Åberg *et al.* (2012) mediated the effects of quetiapine and risperidone on adverse cardiac events respectively. Additionally, Adkins and colleagues (2011) showed that, grouping by medication, risperidone was the antipsychotic with the most number of corresponding significant associations. Clozapine and perphenazine were also significantly associated with a number of different outcomes in this study. The Clark *et al.* GWAS (2013) and the two McClay *et al.* studies (2011a; 2011b) only found significant associations for the four SGAs ziprasidone, risperidone, quetiapine, and olanzapine. Åberg *et al.* (2010) did not indicate drug-specific results.

Despite grouping patients by antipsychotic, the multiphase nature of CATIE meant that the majority of patients were on more than one type of antipsychotic (including clozapine) for the duration of the trial, and patients had not been required to be drug-naïve at the commencement of the study. Although there is an abundance of data available, this trial was not designed with pharmacogenomic application in mind. Nevertheless, CATIE provides a valuable resource of genetic and clinical data, without which the study of antipsychotic pharmacogenomics would be extremely limited. Furthermore, even with smaller, drug-specific GWAS, many SNPs reached genome-wide significance for various treatment outcomes (Table 3.3).

Of the remaining three GWAS cohorts, two stand out as being well-characterised in nature. The first study, by Liou *et al.* (2012), investigated schizophrenia treatment-refractoriness in Han Chinese individuals. Working under the hypothesis that refractoriness may be a “distinct and homogenous subgroup of schizophrenia”, the authors compared refractory individuals to healthy controls. Refractoriness was defined as nonresponse to two antipsychotic trials (chlorpromazine or one of six SGAs) or clozapine. This is in line with recommended treatment-refractoriness criteria (Suzuki *et al.*, 2012). The homogeneity of the cohort,

comparison to controls, and lack of population stratification allowed for greater power to detect associations. Furthermore, replication of three associations in an independent cohort of 273 individuals provided robustness to the results. To evaluate the hypothesis that treatment-refractoriness is a unique endophenotype of schizophrenia, the top SNPs were then genotyped in a larger cohort of schizophrenia cases and healthy controls. Interestingly, none of the SNPs reached significance, providing credence for this concept.

The second well-characterised study was conducted by Malhotra and colleagues (2012). Impressively, three independent cohorts were genotyped to validate the findings of the initial GWAS. This provided a joint P -value of 6×10^{-12} for rs489693, which was consistently associated with poor metabolic outcomes. This is the most significant SNP across all nine GWAS. The significance of this finding is emphasised by the fact that the discovery cohort and one replication cohort were drug-naïve prior to the study. Additionally, the large majority of patients received SGAs, with less than a third of one replication cohort (9% of all individuals) having been administered with haloperidol. Clinical uniformity was further maintained by the use of an exclusively clozapine-administered replication cohort, and exclusion of clozapine-treated patients from the discovery cohort (Malhotra *et al.*, 2012). This GWAS demonstrates that good clinical characterisation is vital, and allows for robust results that provide valuable insight into the genetics of treatment response.

Lastly, in contrast to the Malhotra *et al.* (2012) GWAS, the study designed by Athanasiu and colleagues (2012) has many flaws. The clinical data was obtained in a naturalistic setting, i.e. a hospital. Therefore, antipsychotic type and dose were adjusted when necessary, and patients were on different and multiple types of medication. Although this is how treatment operates in clinical practice, it is not ideal for a pharmacogenomic study attempting to draw statistically sound conclusions about the biology of adverse drug reactions. Patients receiving different classes of psychopharmacological agents (FGAs, SGAs, antidepressants and/ or mood stabilisers) were analysed together, and medications were grouped based on their likelihood of resulting in adverse reactions. This means that different drug types were included in the same subgroup. Bearing in mind that these medication types have different targets, pathways and side effects, finding true genetic associations would be difficult, if not impossible. The authors acknowledge that grouping in such a manner increases heterogeneity and subsequently the risk of type II errors (Athanasiu *et al.*, 2012). A more suitable design would have been to group patients by medication type. Incidentally, 84% of the second group of patients received antipsychotics, and this was the only group for which significant associations were found. Therefore, the results were included in this study, but analyses were interpreted with caution.

3.5.1.2. Significant GWAS findings

As expected, the majority of significant variants were not in protein-coding regions (Table 3.3). For the most part, the authors' annotation and the classification according to SeattleSeq were concordant, however differences were noted for the Clark *et al.* (2013) study. The authors clearly state that rs711355, rs785423, and rs813676 are located in the tight junction protein 1 (*TJP1*) gene, and that rs2980976 occurs in the tumour necrosis factor receptor superfamily, member 11a, NFkB activator (*TNFRFS11A*) gene. However, all four of these SNPs are in fact intergenic and may not affect the function of the latter genes, even though they were interpreted to do so. With regards to regulation, the majority of the GWAS acknowledge that the intergenic findings may affect regulatory processes, but this was not investigated further. For example, Åberg *et al.* (2010) state that the intergenic SNPs rs17022444 and rs7669317 may impact long-range regulatory effects. Considering the SNPs occurring in introns, the GWAS either did not acknowledge that the variant was intronic, or did not discuss the implications thereof, such as potential effects on post-transcriptional processing. For the most part, the nine GWAS follow the trend of interpreting function in terms of the closest gene, and deprioritise genes that have not previously been implicated in antipsychotic response, neurological functioning, or schizophrenia. All of the studies investigated LD to a limited extent, with some performing haplotype analyses, but once again these variants or haplotypes were related back to the closest gene. A more thorough and all-inclusive approach to GWAS interpretation could lead to novel pathway associations and improve biological hypotheses.

3.5.2. Predicted rSNPs and their genomic effects

3.5.2.1. Regions implicated in immunity

Of all the variants accounted for by GWAS and LD analyses, 20 SNPs were classified by RegulomeDB to have regulatory potential (Table 3.5). Interestingly, two of these originate from the Adkins *et al.* (2011) GWAS on metabolic side effects, and the Åberg *et al.* (2012) GWAS on QT interval prolongation, respectively. This reaffirms the importance of accounting for LD structures when interpreting associations, as the most significant rSNPs were not tagged or assessed by previous GWAS. None of the remaining 18 rSNPs have been previously associated with schizophrenia, pharmacogenomics or antipsychotic treatment response. The first GWAS SNP, rs6741819 on chromosome two, was predicted to affect the binding of the transcription factors EBF1, GATA1 and CTCF, and the motif CNOT3. The corresponding PWM for CNOT3 indicates that the position of this SNP is the most conserved relative to other bases, suggesting that substitutions at this locus are undesirable. EBF1 is a

transcription factor that has been shown to play an important role in B cell differentiation within the adaptive immune system (Nechanitzky *et al.*, 2013).

Interestingly, the results include several other regulatory factors involved in immunity. For example, RFX3 is a protein that binds to MHC class II promoters to influence MHC expression (Reith *et al.*, 1995). According to rSNPBase, RFX3 is affected by rs3774959 – the most significant RegulomeDB SNP – and rs230493, both associated via LD with treatment-refractory schizophrenia (Liou *et al.*, 2012). Secondly, CEBPB was predicted to be affected by rs3774933 by both RegulomeDB and rSNPBase analyses. This SNP is similarly in LD with a significant SNP from Liou *et al.* (2012). Lending evidence to this finding, a matched DNase peak for the corresponding C/EBP motif was identified at this locus by RegulomeDB. The CEBPB transcription factor has been shown to play an important role in immune suppression (Marigo *et al.*, 2010). There is a long-standing hypothesis that schizophrenia development is associated with abnormal immune functioning. In fact, the MHC locus is the most replicated genomic region with regards to associations with schizophrenia risk (Sullivan *et al.*, 2012). These results suggest that regulation of different aspects of the immune system could contribute to variation in schizophrenia treatment response, particularly nonresponse. Supporting this idea, a recent meta-analysis of 23 studies revealed that antipsychotics produce anti-inflammatory effects in schizophrenia (Tourjman *et al.*, 2013). Although the exact mechanisms remain unclear, the interplay between antipsychotic response and the immune system should be explored further.

3.5.2.2. Ubiquitous regulatory factors

In addition to the immune system, there are several implicated regulatory factors that are involved in many widespread developmental and regulatory processes. For example, EP300 is a histone acetyltransferase that plays a role in chromatin remodelling and thus regulates gene expression, functioning ubiquitously in processes such as cell growth and proliferation (Ogryzko *et al.*, 1996). Furthermore, TCF4 is a widely expressed transcription factor affected by the rSNPs rs2272676 (associated via LD with treatment-refractoriness) and rs17440909 (associated via LD with patient-rated symptom severity; Clark *et al.*, 2013). Interestingly, *TCF4* has been consistently associated with schizophrenia development in independent studies (Stefansson *et al.*, 2009; Wirgenes *et al.*, 2012; Ripke *et al.*, 2013). Other widespread regulatory effects are illustrated by the RNA-binding proteins associated with post-transcriptional processing (Table 3.6). PABPC1 binds the poly(A)-tail of transcribed mRNA, and has demonstrated involvement in nonsense-mediated decay (Behm-Ansmant *et al.*, 2007). ELAVL1, on the other hand, binds to the 3'-UTR of mRNA and promotes transcript stability (Lebedeva, 2012). SNPs in the regions that bind the latter proteins may

affect general functioning in many cell types, suggesting that pathways nonspecific to drug response may influence antipsychotic treatment outcomes. This notion is supported by the results of the tRap analysis (Tables 3.7 and S1). Considering the three traits used as controls, there are only twelve motifs that are unique to antipsychotic response, and none of these are the top ten most significantly affected motifs.

3.5.2.3. The 4q24 locus and NFKB1

The most notable characteristic of the rSNPs predicted by RegulomeDB and rSNPBase is that results are enriched for the q24 region of chromosome four. Of the 20 SNPs predicted to be significant by RegulomeDB, 14 occur at this locus. Additionally, nine of these scored 1b-f, therefore this region has the most experimental evidence that supports its involvement in regulation, in comparison to other SNPs associated with treatment response. All of the 4q24 rSNPs are in LD with rs230529; one of three SNPs implicated in treatment refractoriness by Liou *et al.* (2012). Interestingly, the original GWAS variant was not predicted by RegulomeDB to affect regulatory regions significantly, even though all the SNPs in this region were in strong linkage disequilibrium for CHBJPT, CHD and JPT+CHB+CHD population groups. This emphasises the importance of analysing each variant within a region instead of interpreting the impact of the tag SNP alone.

As shown in Figure 3.1, the 4q24 region includes the start of transcription of the *NFKB1* gene. The ENCODE tracks in this region show extensive evidence of regulation in different cell lines, including ChIP-Seq and DNase peaks, as well as the histone mark H3K27Ac, which is associated with open chromatin and active transcription. rSNPBase predicted the majority of input SNPs in this region to affect the expression of *NFKB1*. In fact, this gene was predicted to be affected by 104 SNPs – almost half of the rSNPs characterised by rSNPBase. *NFKB1* encodes a highly conserved transcription factor that regulates over 200 genes, and plays important roles in cancer and the immune system (Shishodia and Aggarwal, 2004; Liou *et al.*, 2012). Once again, this implicates immune dysregulation in antipsychotic response. Polymorphisms in *NFKB1* have previously been associated with the pharmacogenetics of anti-tumor necrosis factor (TNF) treatment response (Bank *et al.*, 2014) and with schizophrenia susceptibility (Narayan *et al.*, 2008). Interestingly, a 2015 review identified the most replicated finding in first episode schizophrenia studies to be increased TNF- α levels (Fond *et al.*, 2015). This pro-inflammatory cytokine is regulated by *NFKB1* (Hall *et al.*, 2005), suggesting alterations in this regulatory pathway may be responsible for the identified associations with the disorder.

The enrichment for SNPs that affect expression of *NFKB1* is significant, however one must keep in mind that *NFKB1* may be overrepresented in the results of this study. All significant RegulomeDB SNPs at this locus are proxies for one single SNP. It is possible that the results revealing regulatory potential are indicative of a single signal. Furthermore, variation within or affecting *NFKB1* may not have effects specific to antipsychotic treatment response, since this gene has control over hundreds of downstream loci. Having said that, it is interesting to note that the 4q24 region came up in three of the other original GWAS, which investigated associations with different treatment outcomes (Åberg *et al.*, 2010; Adkins *et al.*, 2011; Clark *et al.*, 2013). This region could be important for treatment response and should be investigated further.

3.5.3. Genes and pathways relevant to antipsychotic response

Apart from *NFKB1*, a number of other affected genes and eQTLs were determined by RegulomeDB and rSNPBase. Notably, the most significant of these have all shown associations with schizophrenia susceptibility or symptom severity in the past. Experimentally, the most significant SNPs are those predicted to act as eQTLs. Overlap between RegulomeDB and rSNPBase was seen for the 4q24 rSNPs and eQTL targets *MANBA* and *COL9A2*. *MANBA* was the third most affected gene in the study and has been linked to schizophrenia (Jungerius *et al.*, 2007). This gene codes for the lysosomal β -mannosidase protein, and mutations can lead to β -mannosidosis (Huynh *et al.*, 2011). The second most affected gene, with a total of 25 rSNPs implicated in its expression (Table 3.9), is the cAMP-specific phosphodiesterase 4D (*PDE4D*) gene, which, like the other two top genes, has also shown links to schizophrenia (Tomppo *et al.*, 2009). The *PDE4D* gene was implicated in the Clark *et al.* (2013) GWAS, and a SNP in one of its introns was significantly associated with patient-rated symptom severity. Interestingly, *PDE4D* inhibition increases dopamine receptor signalling, suggesting that this gene could be a potential antipsychotic target (Halene and Siegel, 2008; Kuroiwa *et al.*, 2011).

Upon first inspection, the pathways and networks predicted by DAVID and GeneMANIA do not indicate any mechanisms that have been traditionally associated with neuropsychiatric disorders or drug response. Relatively few genes were included in each pathway, but mostly high fold change values suggest that these findings are significant. Interestingly, *NFKB1* is a node in three of the four pathways identified by DAVID, demonstrating its diverse and widespread function. Analysis of function with GeneMANIA substantiated the potential role of immunity in treatment response: *NFKB1* and *ACTG1* both function in cell receptor signalling during an immune response (Figure 3.2). Additionally, pointing to effects on neurological functioning, *NFKB1* and *SHC1* are both involved in neurotrophin signalling (Figure 3.3).

Therefore, although these clustered pathways seem unlikely candidates for drug response biology, they should be investigated further to search for clues regarding the mechanisms of antipsychotic treatment outcomes.

3.5.4. Study limitations

The limitations associated with this study concern both the antipsychotic response GWAS and the tools used to analyse them. As discussed, there are relatively few GWAS that have investigated schizophrenia treatment response, and even fewer that have been well-designed and are statistically robust. The majority of results are based on information from one cohort of less than 1000 individuals, therefore the methods applied in this study must be investigated in other samples and population groups. Furthermore, drug-specific cohorts are essential, as mechanisms may differ between FGAs and SGAs, and unique SGAs such as clozapine. There is great need for improvement in clinical characterisation of patients, clinical and genetic homogeneity, and increased GWAS sample sizes. The difficulties in assessing and defining treatment response, as well as obtaining drug-naïve individuals, have thus far restricted these developments.

The range of tools one can use to assess the functional impact of genetic variation is growing rapidly. The analysis of genomic regulation is still in its infancy, and there are many improvements to be made. Inconsistencies between tools and insufficient data may have prevented the discovery of a gene or pathway involved in treatment response. The limited consensus between the tools demonstrates the need for improved bioinformatic design and database curation, as well as the complexity of genetic regulation. Lastly, ENCODE has received criticism for its claim that 80% of the genome is functional (Graur *et al.*, 2013). Although it is the only resource of its kind, one must keep in mind that the data may overestimate functionality within the genome. For example, almost half of the variants analysed were assigned at least one regulatory function by rSNPbase. Although regulation is complex and extensive, such results must be interpreted with caution. With time and additional studies, the regulatory networks behind complex traits and disorders such as antipsychotic response will likely become clearer.

3.6. Conclusion

This study has provided a novel approach to the functional analysis of genomic variation. With a bioinformatics pipeline that can be applied to other complex traits, regulatory changes and downstream effects can elucidate the mechanisms involved in the workings of antipsychotics and the development of adverse drug reactions. Given that previous studies

have revealed little about the biology of antipsychotics, a hypothesis-free approach such as this one provides the best means of understanding the heterogeneity of treatment response.

Importantly, this study reaffirmed the functionality of noncoding regions of the genome, and the widespread impact that rSNPs can have; affecting proximal, long-range and post-transcriptional processes. The most significant finding of this study was the predicted role of *NFKB1* in treatment-refractory schizophrenia, which provided further evidence that refractoriness is a distinct endophenotype of the disorder. The 4q24 locus may be important for immune-mediated response to antipsychotics, and it is essential that this is investigated in future. Additionally, this study lends evidence to the hypothesis that there is extensive overlap between schizophrenia risk and treatment response pathways, perhaps particularly with regards to immune functioning. Much like each symptom domain of schizophrenia, it is likely that different treatment outcomes reflect different pathways, and that cumulative variants in both coding and noncoding regions contribute to their inception and severity.

Future research on antipsychotic response should involve large-scale GWAS coupled with bioinformatic and functional analyses. Ideally, results should be validated in multiple cohorts, and include previously underrepresented populations, since these groups experience the majority of the mental health burden. This process begins with better study design and clinical assessments and, with the use of bioinformatics, has the potential to conclude with improvements in our understanding of complex disorders. This paves the way for better treatment response and medication tailored for each individual.

PART 2

Association analyses

CHAPTER 4:**Associations between predicted regulatory variants and antipsychotic treatment outcomes in a South African schizophrenia cohort****4.1. Summary**

Although antipsychotics are effective at reducing the positive symptoms associated with schizophrenia, remission is currently not a realistic goal for the majority of patients, and relapse is often a reality. Pharmacogenomics has the potential to elucidate the genetic factors contributing to this heritable trait, and ultimately guide future drug design and treatment improvement. This study attempted to validate the findings of a novel bioinformatic pipeline that assessed previous antipsychotic pharmacogenomic GWAS hits for regulatory impact on treatment response. TaqMan® OpenArray® was utilised to genotype 31 SNPs in a South African cohort, which were subsequently assessed for associations with treatment outcomes and ADRs with the use of mixed-effects repeated measures analyses. Significant associations were observed for several outcomes, including changes in PANSS scores, refractoriness, remission, and metabolic side effects.

The associations that survived Bonferroni correction were located on chromosome 4q24, which was the most significant regulatory locus identified in the bioinformatic portion of the study. Furthermore, this region has been previously implicated in schizophrenia. The SNPs rs230493 and rs3774959 were both significantly associated with greater post-treatment PANSS Negative scores ($P < 0.00001$). Additionally, a 14-SNP haplotype containing these two variants was predicted to elicit a 4.41% higher post-treatment positive symptom score. These results validate the importance of the 4q24 region in antipsychotic response, the biological overlap of schizophrenia susceptibility and drug response, and the hypothesised role of genomic dysregulation in adverse treatment outcomes. Interestingly, nonconcordance was observed with regards to dichotomous outcomes of refractoriness and remission, and significant variants. This suggests that clinical characterisation of these states should be re-evaluated with respect to genetic variables. Lastly, the study of a well-characterised South African cohort not only contributes to our knowledge of pharmacogenomics, but also improves our understanding of this genetically rich and distinct population group. These findings have the potential to aid in improving treatment outcome in schizophrenia.

4.2. Introduction

Schizophrenia is a severe, lifelong psychiatric disorder for which treatment is often complex and ongoing. Treatment response can be measured by several different scales that assess changes in symptom severity. This includes the PANSS, which accounts for different symptom domains, including positive and negative symptoms (Kay *et al.*, 1987). Additionally, there are a few terms that are important to define when assessing treatment response. Firstly, remission has been classified by the Remission in Schizophrenia Working Group as a period of at least six months in which symptoms are absent or mild, and do not affect an individual's functioning (Andreasen *et al.*, 2005). In contrast, treatment-refractoriness or nonresponse is generally established when a patient's symptoms do not adequately improve after at least two trials of different antipsychotics (Suzuki *et al.*, 2012). It is important to note that these states are not absolutes and are difficult to define and predict, due to differences between individuals and different response criteria. One reliable predictor of long-term outcome is early treatment response. Studies have shown that an individual responding well as early as two weeks into treatment is an indicator of later remission, whilst refractoriness can be predicted by early nonresponse (Kinon *et al.*, 2008; Chiliza *et al.*, 2015a).

The complexity and heterogeneity of treatment response is largely brought about by the genetic differences between individuals in drug metabolism, neurotransmitter, and other pathways (Blanc *et al.*, 2010; Klein and Zanger, 2013; Ni *et al.*, 2013). However, studies to date have mostly provided inconsistent results. Much like other complex traits, there are likely to be hundreds to thousands of common variants across the genome that cumulatively contribute to individual treatment response phenotypes (Arranz and de Leon, 2007). With large and well-characterised sample groups, pharmacogenomics enables the discovery of these variants. Specifically, GWAS have recently been employed to study antipsychotic response. While this approach trumps *a priori* candidate gene studies by analysing variants across the genome, the majority of GWAS lack sufficient biological interpretation. Often, variants are considered in isolation and exclusively with regards to the function of their neighbouring gene (Åberg *et al.*, 2012; Clark *et al.*, 2013). This restricts the creation of new hypotheses and further understanding of treatment response mechanisms.

Overall, GWAS have neglected individuals of non-European descent. This potentially precludes the development of effective treatment for all individuals, since populations vary extensively with regards to genetics. Indeed, the allele frequency of a GWAS variant has been shown to vary 40-fold between different ethnicities (Adeyemo and Rotimi, 2010). Africans have been shown to be the most genetically diverse population group, yet they are

underrepresented in genetic studies (Drögemöller *et al.*, 2011). This group's unique genetic make-up could provide invaluable information about variation contributing to complex phenotypes such as treatment response. Further motivation to study African populations is the fact that they suffer a greater mental health burden due to poor resources and public health care (refer to chapter one).

The current study combines the need to better interpret GWAS results, while also considering previously understudied population groups. To improve functional interpretation of GWAS, a previous study employed a bioinformatic pipeline to assess all variants associated with treatment response by GWAS (chapter two). This included making use of several tools and large-scale datasets to assess the noncoding variants' regulatory potential, as well as the impact of significant coding variants. The study found merit in approaching GWAS interpretation more comprehensively, as results suggested the novel hypothesis that *NFKB1* and the immune system may contribute to antipsychotic response. However, the complexity of treatment response and the purely bioinformatic nature of this study require these results to be validated. Therefore, the current study aimed to perform association analyses in a well-characterised South African cohort of FES patients. As a second phase to the bioinformatic pipeline, this could provide validation for new treatment response hypotheses, as well as decrease the research gap between different population groups. The coupling of bioinformatics with specific studies in different ethnicities provides a comprehensive method to elucidate common genetic factors contributing to antipsychotic treatment outcomes. In the long term, understanding the biological mechanisms of antipsychotics provides a platform for better drug design and effective treatment of schizophrenia.

4.3. Materials and methods

Refer to Electronic Sources (p. 107) for dates of access to online tools.

4.3.1. Patient samples

A South African FES cohort of 103 patients (median age 23 ± 7 years; 74% male) was used to investigate associations with SNPs previously implicated in treatment response. The cohort consisted of 82 SAC, 13 Xhosa and 8 Caucasian individuals. Patients were recruited over four years at Stikland Hospital in the Western Cape and assessed with the Structured Clinical Interview for the DSM-IV (American Psychiatric Association, 1994). Demographic and medical data were obtained at the time of recruitment. Signed written and informed consent was provided by all patients or their caregivers prior to the study. Ethical approval

was obtained from the Human Research and Ethics Committee (HREC), Faculty of Health Sciences, Stellenbosch University (ethics numbers for clinical and genetic aspects: N06/08/148 and 1907/005 respectively).

All patients received treatment with flupenthixol decanoate, an FGA, by injection. Dose was gradually increased until remission was achieved – defined according to the Remission in Schizophrenia Working Group (Andreasen *et al.*, 2005) – or until the maximum recommended dose was reached. Response to treatment was measured by the PANSS over a period of 12 months, with measurements taken every two weeks for the first six weeks, and every three months thereafter. A decrease of 25% or greater in PANSS scores at six weeks was considered indicative of early response to treatment in this study. Furthermore, treatment-refractory patients were defined as those who 1) discontinued treatment because of poor response, 2) showed a < 25% reduction in total PANSS scores at 12 months, or 3) had a PANSS score > 70 at 12 months (Chiliza *et al.*, 2015a). The latter two categories of patients must have completed treatment for at least three months without relapse in order to qualify as treatment-refractory. Lastly, data on metabolic outcomes was recorded: BMI, lipid profiles, and changes in cholesterol were measured for each patient at three, six, nine and twelve months (Chiliza *et al.*, 2015b).

Prior to the current study, genomic DNA (gDNA) was extracted from whole blood samples from each patient, using the Miller *et al.* (1988) protocol.

4.3.2. SNP prioritisation

Variants previously predicted to affect regulation were prioritised for genotyping in the cohort (refer to chapter three). Regulatory predictions include those from RegulomeDB (<http://regulome.stanford.edu/>), rSNPBase (<http://rsnp.psych.ac.cn/>) and sTRAP (http://trap.molgen.mpg.de/cgi-bin/trap_two_seq_form.cgi). SNPs were prioritised according to the RegulomeDB results. Even though this includes SNPs with a score > 3, these findings are the most robust compared to other tools, since they are based on experimentally validated data such as ENCODE (Boyle *et al.*, 2012). In contrast, TRAP works only with computational predictions (Manke *et al.*, 2010), and, although it deemed 222 SNPs to be involved in regulation, rSNPBase does not use a scoring system to rank variants. There is, however, extensive overlap between the top RegulomeDB variants and rSNPs predicted by rSNPBase. RegulomeDB variants were coupled with their proxy SNPs from antipsychotic response GWAS (provided the GWAS SNP was not already classified as an rSNP) for a set of 30 variants. Additionally, one exonic SNP (rs17727261) was predicted to be nonsynonymous by PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>). This variant was

also included for genotyping, creating a total of 31 SNPs for further analysis. The final set of genotyped SNPs, along with their associated response traits, is indicated in Table 4.1. In cases where a SNP failed assay design (4.3.3 below), the variant and its GWAS partner were replaced by the next most significant pair according to RegulomeDB.

4.3.3. SNP genotyping

The 31 SNPs in Table 4.1 were genotyped in the South African FES cohort with the use of TaqMan® OpenArray® Real-Time Polymerase Chain Reaction (PCR) (Life Technologies™, New York, USA). TaqMan® assays were obtained from the SNP Genotyping Assay Search Tool (<http://www.lifetechnologies.com/za/en/home/life-science/pcr/real-time-pcr/real-time-pcr-assays/snp-genotyping-taqman-assays.html>). In cases where no predesigned assay was available, a custom assay was designed by Life Technologies™ in New York. SNPs that failed custom assay design or functional testing were excluded and replaced. The customised 31-SNP assay was manufactured by Life Technologies™ and shipped directly to the University of Utah DNA Sequencing and Genomics Core Facility for genotyping, along with genotyping master mix, OpenArray® 384-Well Sample Plates and necessary consumables. The SNPs and their corresponding assays are indicated in Table S3.

Concentrations of previously extracted DNA samples for all 103 patients were measured using the NanoDrop spectrophotometer (NanoDrop® ND-100, NanoDrop Technologies Inc., Wilmington, Delaware, USA) and diluted to 25 µl at 80 ng/µl per sample. Thereafter, 20 µl of each sample was added to two MicroAmp® 96-well plates (Applied Biosystems™, California, USA). Duplicate samples were included as positive genotyping controls, and two empty wells per plate served as negative controls. The plates were sealed with optical adhesive film, frozen, and shipped on dry ice to the University of Utah. Subsequently, genotyping was performed according to the manufacturer's instructions and analysed with the OpenArray® SNP Genotyping Analysis Software version 1.3.1.

Table 4.1: SNPs genotyped in the South African FES cohort, including predicted rSNPs and corresponding GWAS SNPs.

	Position (GRCh37/hg19)	SNP	Relevance ^a				Associated response measurement	
			RegulomeDB Score	rSNPBase	TRAP	GWAS		PP-2
Chr 1	70921172	rs10458561	3a			x	QTc interval prolongation	Åberg <i>et al.</i> , 2012
	160077853	rs6688363	4	x		x	Clinical impression of severity (CGI-S)	Clark <i>et al.</i> , 2013
	160626060	rs10218843	N/A			x	Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	160630142	rs11265461	N/A			x	Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	160634587	rs6427540	4	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	236313126	rs7520258	N/A	x		x	Neurocognition: working memory	McClay <i>et al.</i> , 2011a
Chr 2	7147972	rs6741819	2b	x		x	Triglycerides	Adkins <i>et al.</i> , 2011
	125281909	rs17727261	N/A	x		x	Negative symptoms (PANSS)	McClay <i>et al.</i> , 2011b
	139259221	rs62161711	4	x			Patient global impression (PGI) scale	Clark <i>et al.</i> , 2013
	139278921	rs10170310	N/A	x	x	x	Patient global impression (PGI) scale	Clark <i>et al.</i> , 2013
Chr 4	103414174	rs747559	1f	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103426338	rs3774933	1b	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103443568	rs1599961	1f	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103449040	rs230534	4	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103450166	rs230532	1f	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103457417	rs230529	N/A	x		x	Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103458824	rs230526	2b	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103463006	rs118882	4	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103465611	rs230520	1f	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103481350	rs230505	1d	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103481560	rs230504	1f	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103485779	rs230492	4	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103486215	rs230493	1f	x	x		Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103487299	rs230495	3a	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
	103495531	rs230539	4	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012
103511113	rs3774959	1b	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012	
103515312	rs4648055	1f	x			Treatment-refractory schizophrenia	Liou <i>et al.</i> , 2012	
Chr 5	58957830	rs17440909	3a	x			Patient impression of severity (PGI)	Clark <i>et al.</i> , 2013
	58999041	rs17742120	N/A	x	x	x	Patient impression of severity (PGI)	Clark <i>et al.</i> , 2013
Chr 8	89566902	rs1352318	2c				BMI	AthanasIU <i>et al.</i> , 2012
Chr 12	31357101	rs10492354	1f		x		Negative symptoms (PANSS)	McClay <i>et al.</i> , 2011b

^a Relevance of SNP with regards to regulatory evidence (RegulomeDB, rSNPBase, TRAP) or protein impact (PolyPhen-2), as well as proxy GWAS SNPs. PP-2 = PolyPhen-2; Chr = chromosome; N/A = not applicable.

4.3.4. Statistical analyses

Allele and genotype frequencies for successfully genotyped SNPs were determined, and deviations from Hardy-Weinberg equilibrium (HWE) were calculated by means of a Pearson's Chi-square (χ^2) test or analogue to Fisher's Exact Test with SNPStats (Solé *et al.*, 2006; <http://bioinfo.iconcologia.net/snpstats/start.htm>). SNPs with $P < 0.01$ were considered to deviate from HWE. SNP frequencies within the cohort were compared to those in HapMap (The International HapMap Consortium, 2003) and 1000 Genomes (The 1000 Genomes Project Consortium, 2010) population groups. Subsequently, LD between SNPs was assessed with Haploview version 4.2 (Barrett *et al.*, 2005). Haploview designated haplotype blocks by assessing pairwise LD. SNPs with $r^2 \geq 0.8$ were considered to be in LD, but D' confidence intervals ($D' > 0.7 - > 0.98$) were also assessed for comparison (Gabriel *et al.*, 2002). The haplotypes were subsequently assessed for associations with treatment outcomes, along with all individual SNPs.

Allelic, genotypic and haplotypic association analyses were performed in the R Linear and Nonlinear Mixed Effects Models package (Pineiro *et al.*, 2014). To determine the presence of associations with treatment outcomes (measured by changes in PANSS scores as well as several metabolic variables), mixed-effects model repeated measures analyses were conducted. Additionally, any associations with dichotomous outcomes, i.e. treatment-refractoriness, remission, or early response, were determined with logistic regression models.

Since the SAC population is highly admixed, any spurious associations due to population stratification were accounted for by correcting for ancestry contributions. This was accomplished by utilising ancestry informative markers (AIMs) as covariates to estimate ancestry proportions in ADMIXTURE (Alexander *et al.*, 2010; Daya *et al.*, 2013). The ancestry proportions of the SAC individuals are indicated in Figure 4.1. In addition to proportion ancestry, all analyses were adjusted for age, gender, and ethnic group, and the mixed-effects model repeated measures analysis was adjusted for baseline PANSS scores when assessing change in PANSS over time.

Bonferroni was used to correct for multiple testing after association analyses. Modes of effect and inheritance were tested for the most significant associations, and estimates of effect size with 95% confidence were determined for all models.

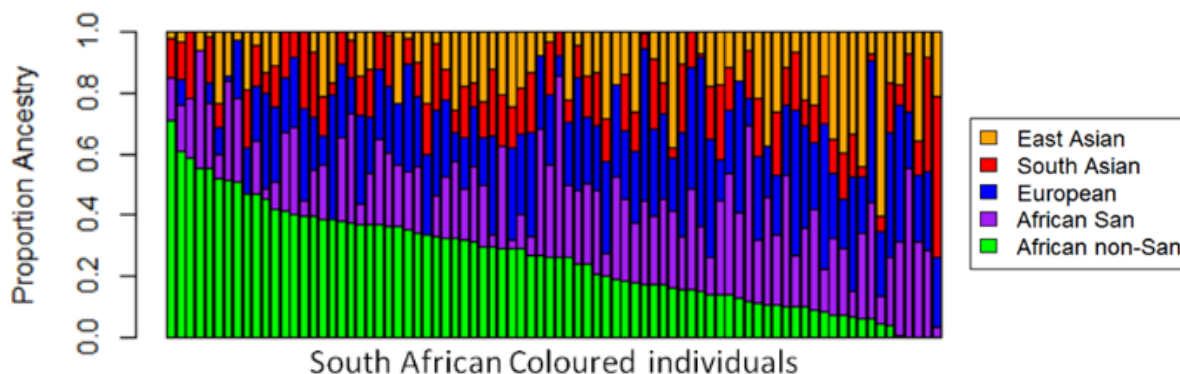


Figure 4.1: Ancestry contributions from five populations in the SAC FES individuals (Drögemöller, 2013).

4.4. Results

4.4.1. Clinical outcomes

The FES cohort was assessed for various treatment response outcomes. Firstly, considering the PANSS scores and criteria defined by the Remission in Schizophrenia Working Group, 77 patients (74.8%) achieved early response at six weeks, and 10 individuals (9.7%) were classified as being treatment refractory. 58 patients (56.3%) achieved full remission by the end of the twelve month treatment period (Chiliza *et al.*, 2015a). With the use of linear mixed effect models for continuous repeated measures, Chiliza and colleagues (2015b) observed significant weight gain within the cohort ($P < 0.0001$), with 58.2% of individuals gaining more than 7% weight. Additionally, increases in BMI ($P < 0.0001$) and triglycerides ($P = 0.03$), and a significant decrease in HDL cholesterol ($P = 0.005$) were observed.

4.4.2. SNP genotyping

All 31 SNPs were successfully genotyped, with an average call rate of 97.7%. Sample duplicates (FS027, FS095, and FS122) displayed concordant genotypes across all assays, except in a few cases where one or both failed to amplify, where the genotype was classed as “undetermined”. The allelic discrimination plot for rs6427540 is shown in Figure 4.2. For this SNP, all genotypes were determined, and the assay achieved a call rate of 100%.

All genotyped SNPs demonstrated a MAF ≥ 0.05 in the cohort of 103 patients, and are thus considered common variants within this cohort (Iyegbe *et al.*, 2014). Additionally, all SNPs were in HWE ($P \geq 0.01$). The SNP with the lowest frequency, rs17727261 (MAF = 0.05) has similar frequencies in both HapMap and 1000 Genomes CEU samples, but does not occur in the other population groups, as shown in Figure 4.3. On the other hand, the SNP with the highest MAF (0.49), rs230505, is the major allele in all other HapMap and 1000 Genomes populations. Population comparisons for two other SNPs of intermediate frequency in the

South African cohort are also shown in Figure 4.3. Overall, the majority of SNPs displayed frequencies similar to the combined average of other population frequencies.

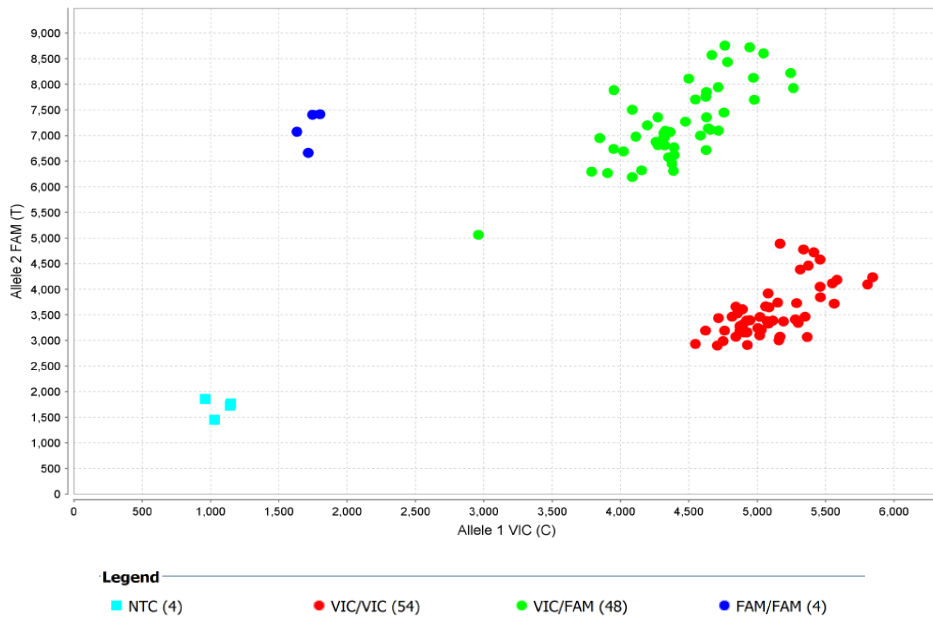


Figure 4.2: Allelic discrimination plot for rs6427540. VIC® and FAM® relative dye intensities indicate genotype: VIC/VIC (CC); VIC/FAM (CT); FAM/FAM (TT). NTC = no template control.

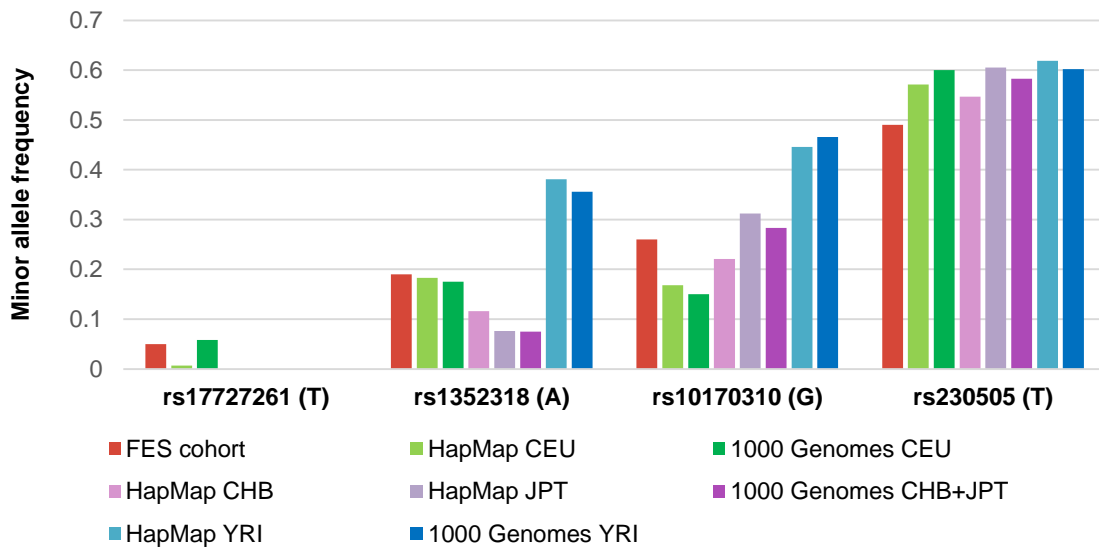


Figure 4.3: Frequency comparisons between the FES cohort and HapMap and 1000 Genomes populations. Allele depicted refers to the minor allele within the FES cohort.

4.4.3. Haplotype analyses

Analysis with Haploview showed that SNPs on chromosome four are in strong LD within the FES cohort, similar to the LD observed in previous analysis of a Han Chinese cohort (Liou *et al.*, 2012). This region, with LD measured by r^2 , is indicated in Figure 4.4. The same haplotype blocks were designated according to D' confidence intervals, shown in Figure S1. The two designated haplotype blocks for chromosome four were analysed further for associations with treatment outcomes. Similarly, haplotypes were identified for variants on chromosomes one, two, and five. All haplotypes and their inferred frequencies within the FES cohort are indicated in Tables S4a-e. Only those with frequencies greater than 0.01 are shown, and haplotypes with lower frequencies were removed from further analyses.

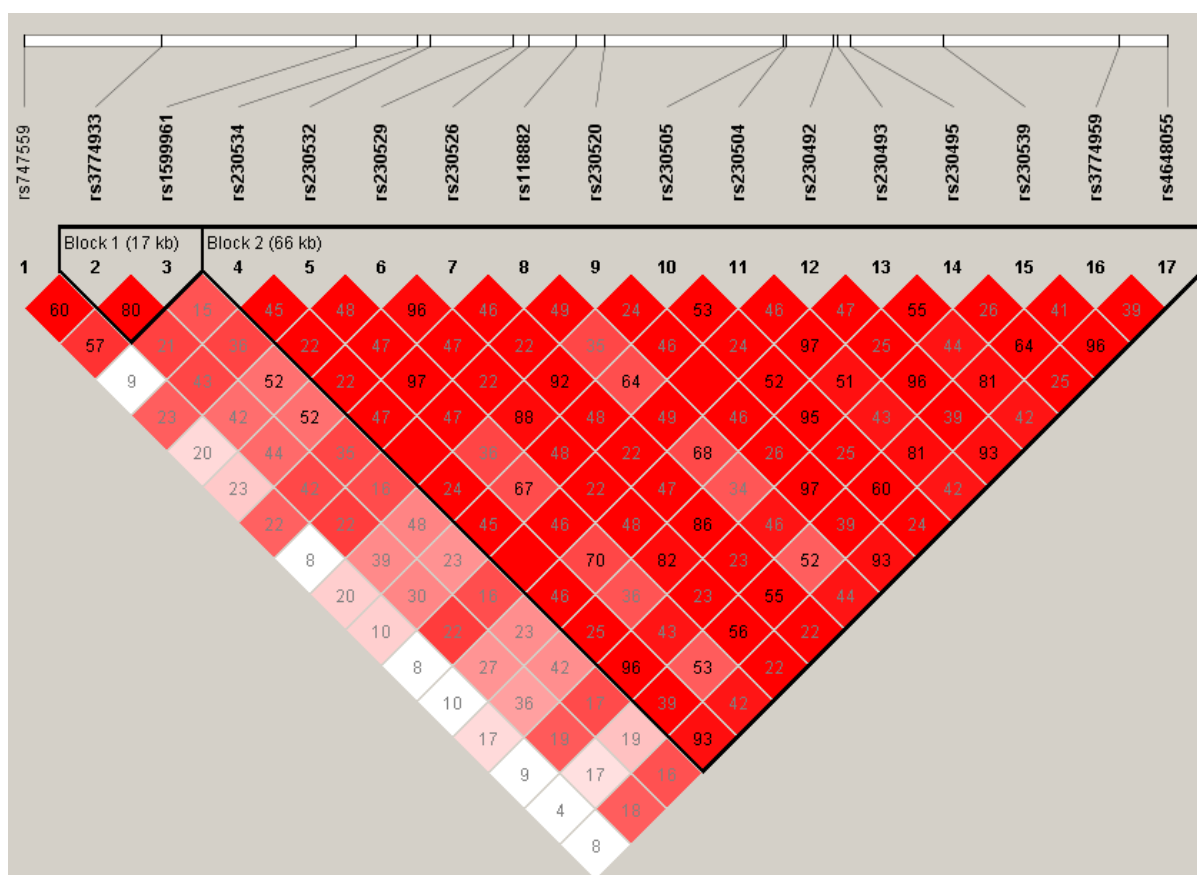


Figure 4.4: Two haplotype blocks on chromosome four, designated by Haploview version 4.2 ($r^2 \geq 0.8$; $\text{LOD} \geq 3$) (Barrett *et al.*, 2005). Dark red squares indicate significant LD between SNPs; r^2 values are shown as a percentage in each square.

4.4.4. Association analyses

In total, there were 23 SNPs and 10 haplotypes that were significantly associated with treatment outcomes ($P < 0.05$), and 97 unique associations, as shown in Table S5. Three associations survived Bonferroni correction and are highlighted in the table. Table 4.2 shows

Table 4.2: Top significant SNP and haplotype associations with treatment outcomes in the FES cohort, with effect models and sizes indicated.

Chr	Variant/ haplotype	Associated response measurement	Effect model	Comparison	P-value	Effect ^a	95% CI	
4	rs230504	PANSS Negative	Dominant	TT + CT vs. CC	0.0001	1.47	0.71	2.24
	rs230493	PANSS Negative	Genotype	TA vs. TT	< 0.0001	1.98	1.20	2.76
	rs230493	PANSS Negative	Genotype	AA vs. TT	< 0.0001	0.38	-0.84	1.62
	rs230493	PANSS Total	Dominant	TA + AA vs. TT	0.0004	1.28	0.58	1.98
	rs230495	PANSS Negative	Dominant	AA + AG vs. GG	0.0006	1.59	0.68	2.50
	rs3774959	PANSS Negative	Genotype	GA vs. GG	< 0.0001	1.82	1.01	2.63
	rs3774959	PANSS Negative	Genotype	AA vs. GG	< 0.0001	0.31	-0.85	1.48
	rs3774959	PANSS General	Genotype	GA vs. GG	0.0004	1.58	0.78	2.39
	rs3774959	PANSS General	Genotype	AA vs. GG	0.0004	0.57	-0.59	1.75
	rs3774959	PANSS Total	Genotype	GA vs. GG	0.0001	1.67	0.91	2.44
	rs3774959	PANSS Total	Genotype	AA vs. GG	0.0001	0.72	-0.39	1.84
	C.A.T.A.C.A.G.T.G.A.A.A.A.G*	PANSS Positive	Each additional haplotype		0.0002	4.41	2.10	6.78

^a Effect measured in percentage change in PANSS score per month. Shaded variants/ haplotypes maintained significance after correcting for multiple testing. CI = confidence interval.

* SNPs: rs230534, rs230532, rs230529, rs230526, rs118882, rs230520, rs230505, rs230504, rs230492, rs230493, rs230495, rs230539, rs3774959, rs4648055.

the top five significant SNPs and haplotypes, including those that survived correction for multiple testing, together with their associated outcomes, inheritance models, and predicted effect sizes with confidence intervals. As shown in the table, change in PANSS scores – particularly PANSS Negative scores – was the most significant treatment outcome within the cohort. Surviving associations are limited to regions on chromosome four, with the significant haplotype containing the minor alleles of the two individually significant SNPs (rs230493 and rs3774959; Table 4.2).

4.5. Discussion

This study made use of the results of a novel bioinformatics pipeline designed for more comprehensive and biologically relevant interpretations of significant GWAS variants. These variants have been previously implicated (either by GWAS or LD and predicted regulatory effects) in antipsychotic treatment outcome. The SNPs and relevant haplotypes were tested for associations with various schizophrenia treatment outcomes, including improvement or decline in symptoms, changes in BMI and other metabolic outcomes, as well as indicators of early response or nonresponse to treatment.

4.5.1. Clinical outcomes

General treatment outcomes were assessed based on changes in PANSS scores over a 12 month period. Firstly, 56% of the FES cohort achieved remission, i.e. their symptoms were deemed absent to mild for at least six consecutive months (Andreasen *et al.*, 2005; Chiliza *et al.*, 2015a). A majority of individuals achieving remission is expected, as first-episode patients generally respond well to treatment in comparison to second- or multi-episode patients (Kahn and Sommer, 2015). It is well-known that a shorter DUP is linked to an improved response (Perkins *et al.*, 2004; Jeppesen *et al.*, 2008). Secondly, only 10% of patients were classified as treatment-refractory, according to the criteria outlined in 4.3.1. Again, refractoriness is more common in multi-episode patients, and it has been suggested that nonresponsiveness often emerges in individuals that have previously responded to medication and subsequently relapsed (Caspi *et al.*, 2004). Lastly, a high percentage of patients achieved early response, i.e. their total PANSS scores decreased by at least 25% after six weeks of treatment. Antipsychotics bring about the greatest symptom changes within the first few weeks of use (Agid *et al.*, 2003). Since the majority of patients achieved remission, and early symptom improvement is an accurate predictor of remission, this high percentage is expected (Kinin *et al.*, 2010).

Although the majority of the cohort achieved remission, severe adverse metabolic outcomes were observed. Over half of the cohort gained > 7% weight, which was accompanied by a significant increase in BMI and triglycerides, and a significant decrease in HDL cholesterol – all risk factors for metabolic syndrome (Chiliza *et al.*, 2015b). Such metabolic reactions are common to SGA-administered patients (Newcomer, 2005; Tandon *et al.*, 2010). The distinguishing characteristic of SGAs is that their improved design does not lead to adverse motor effects as seen with FGA usage, but can cause undesirable metabolic reactions (Brennan, 2014). Nevertheless, FGAs, particularly those of low potency such as flupenthixol, have also been associated with metabolic ADRs to a lesser extent (Leucht *et al.*, 2009). The severity of metabolic outcomes in the FES cohort can be further explained by the observation that treatment-naïve first-episode patients are particularly sensitive to antipsychotics (McEvoy *et al.*, 1991; Oosthuizen *et al.*, 2004). In fact, many previous studies have shown considerable adipogenic side effects in FES treatment cohorts (Strassnig *et al.*, 2007; Tarricone *et al.*, 2010; Correll *et al.*, 2011). The clinical data suggests that improving – or ideally, preventing – metabolic side effects should be a priority for FES patients, whether they are receiving FGAs or SGAs.

4.5.2. SNP genotyping and frequency comparisons

All 31 SNPs were informative and could be analysed further for associations with treatment outcomes. When compared to HapMap and 1000 Genomes population frequencies, the majority of variants displayed frequencies in the cohort that were intermediate between other populations groups, for example, rs1352318 and rs10170310, shown in Figure 4.3. This is expected, since 80% of the cohort comprises SAC individuals, a population group that is highly admixed and contains ancestry contributions from several different populations documented on HapMap and 1000 Genomes (Daya *et al.*, 2013).

4.5.3. Associations with treatment outcomes

Upon first inspection of the findings from the association analyses, it is clear that there are many SNPs significantly correlated with changes in PANSS scores. In fact, all but nine of the 31 variants are nominally associated ($P < 0.05$) with at least one treatment outcome (Table S5). This is not unexpected, since these SNPs were chosen for genotyping on the basis of previous associations with treatment outcomes. The majority of these SNPs occur on chromosome four and have previously been associated with treatment-refractoriness (Liou *et al.*, 2012), and several SNPs are associated with more than one outcome. For example, rs3774959 reached significance for changes in all PANSS symptom domains, and a large haplotype containing this SNP was also significantly associated with early treatment response and increases in triglyceride levels. These associations suggest that this locus

plays an important role in antipsychotic response. Furthermore, the rs3774959 variant was the top result in previous regulatory analyses, classified as an eQTL by RegulomeDB and rSNPBase (Tables 3.5 and 3.6). Therefore, this variant could contribute to dysregulation of one or more gene targets that are involved in drug response. In particular, this variant was predicted to affect expression of *MANBA* – previously implicated in schizophrenia – as well as the binding of RFX3, a protein involved in immunity. The region containing rs3774959 and other significantly associated variants – particularly those SNPs and haplotypes that survived Bonferroni correction – will be discussed further.

4.5.3.1 The 4q24 region

The large haplotype on chromosome four that contains the minor alleles for rs230504, rs230493, rs230495, and rs3774959 was associated with PANSS Positive, Negative, and Total scores, and maintained significance with positive scores after correction for multiple testing (Tables 4.2 and S5). Many of the 14 SNPs within this haplotype were also nominally associated with at least one PANSS domain, although only the associations of rs230493 and rs3774959 with negative symptoms survived Bonferroni correction. These results suggest that the four minor alleles of the haplotype produce an additive effect when combined, in relation to positive symptom changes over the course of treatment. Although this haplotype occurred at a low frequency in the cohort (0.03), its presence is associated with a substantial difference in change in PANSS positive scores (4.41% per month higher; Table 4.2). The effect size here is notable, given that the two individually significant SNPs demonstrated much smaller effects, with the lower and upper confidence intervals for their homozygous (AA) genotypes on either side of zero. Therefore, this haplotype should be further studied in larger as well as replicative cohorts to validate these findings.

The 4q24 rSNPs originate from the Liou *et al.* (2012) study or are in LD with the SNPs identified in this GWAS. The authors found that rs230529, the tag SNP of this region, was associated with treatment-refractory schizophrenia. Direct comparison between the Han Chinese cohort studied by Liou *et al.* (2012) and the South African cohort is difficult for a number of reasons. Firstly, only 10 patients in the FES cohort were classified to be treatment-refractory, since the majority of patients achieved remission (Chiliza *et al.*, 2015a). This limits the statistical power of the analyses and the ability to identify significant associations. In fact, none of the 4q24 SNPs were associated with treatment-refractoriness as defined in the cohort. Secondly, the clinical environment was vastly different and refractoriness was defined in different ways. Liou *et al.* (2012) characterised treatment-refractoriness as two failed antipsychotic trials (chlorpromazine or an SGA) or nonresponse to clozapine. The Severity (CGI-S) and Improvement (CGI-I) subscales of the Clinical Global

Impression scale were used to determine the extent of treatment response (Conley and Kelly, 2001). On the other hand, the FES cohort was only treated with an FGA, and the PANSS scale was used to determine response or nonresponse. These differences in assessment make direct comparison problematic. Nevertheless, the top associations within the FES cohort were all predicted to worsen PANSS score outcomes, with effect sizes ranging from 0.31 to 1.98% per month for individual SNPs. This relative lack of improvement points to nonresponse for these symptom domains, which is consistent with the association observed by Liou *et al.* (2012), and may contribute to treatment-refractoriness.

The symptom domain most commonly associated with this region in the FES cohort is the PANSS Negative subscale. Two SNPs in the haplotype were highly significant for post-treatment increases in PANSS Negative scores ($P < 0.00001$). Interestingly, the region significantly associated with worsened negative symptoms within the FES cohort was associated with refractoriness by Liou and colleagues (2012). The negative symptoms of schizophrenia, i.e. avolition and blunted emotion, are especially complex and difficult to treat (Millan *et al.*, 2014). These symptoms tend to linger even when positive symptoms have improved, and have been shown to influence the extent of residual cognitive deficits and functional outcomes in schizophrenia patients (Lin *et al.*, 2013; Malaspina *et al.*, 2014; Woodward *et al.*, 2014). The associations with PANSS Negative scores unveiled in this study suggest that i) there is a genetic link between severe, persistent negative symptoms and treatment-refractoriness, ii) there are regulatory mechanisms involved in the pathophysiology and manifestation of negative symptoms and perhaps cognitive deficits in the disorder, and iii) these novel loci could be used as new drug targets to improve negative symptoms and prevent treatment-refractoriness in schizophrenia. Lastly, one must keep in mind the hypothesis that treatment-refractoriness is a biologically distinct endophenotype of schizophrenia (Liou *et al.*, 2012), which is supported by the findings of the bioinformatic analyses. The involvement of the 4q24 region may point to a unique mechanism that influences the progression of treatment nonresponse.

This region of variation is found within *NFKB1* and was previously predicted to affect its regulation (see chapter three). *NFKB1*, in turn, regulates hundreds of genes, and has been linked to immune functioning (Shishodia and Aggarwal, 2004), schizophrenia susceptibility (Narayan *et al.*, 2008), and anti-TNF treatment response (Bank *et al.*, 2014). Additionally, the highly significant 4q24 SNPs, rs230493 and rs3774959, are eQTLs for several genes, including *CACNB1* and *MANBA* (Table 3.5). *CACNB1* encodes a neuronal calcium channel subunit, and decreased expression of this gene has been implicated in schizophrenia (Smolin *et al.*, 2012). As discussed in chapter three, variation in the *MANBA* gene has also

been associated with schizophrenia risk (Jungerius *et al.*, 2007). The significant associations within the FES cohort validate the findings of the bioinformatic pipeline, and add weight to hypotheses regarding a connection between treatment response and immunity, as well as a biological overlap between schizophrenia risk and treatment outcomes.

4.5.3.2 Refractoriness, remission, and early response

This study revealed four associations for the dichotomous outcomes, i.e. treatment-refractoriness, remission, and early treatment response, however none survived correction for multiple testing. Firstly, two variants, rs7520258 and rs1352318, were associated with treatment-refractoriness in the cohort. Interestingly, neither of these variants were associated with any other treatment outcomes, such as changes in PANSS scores. Again, this could point to a unique mechanism involved in refractoriness. The rs7520258 variant was previously associated with neurocognition outcomes (McClay *et al.*, 2011a), and rs1352318 was associated with changes in BMI (Athanasiu *et al.*, 2012). Both variants were predicted by bioinformatic analyses to play a regulatory role (Table 4.1), although the biological consequences of these rSNPs remain unclear.

There are no variants in Table 4.2 that were associated with increased or decreased PANSS scores, as well as refractoriness or remission, respectively. Since these states of response are classified according to PANSS score changes, one would expect commonalities between variants for the continuous and dichotomous outcomes. On the contrary, instances of nonconcordance were observed. As discussed, rs1352318 was nominally associated with refractoriness ($P = 0.0338$), but unexpectedly showed an association with remission ($P = 0.0083$). Furthermore, a chromosome four haplotype containing minor alleles for rs230495 and rs3774959 was significantly associated with early response at six weeks *and* worsened endpoint response. As discussed, these two SNPs – individually and in combination – correlate with poorer PANSS outcomes, which is incongruent with their involvement in early response. Since there are no variants that were associated with both early treatment response and improved outcome or remission, these loci do not serve as predictive biomarkers of treatment outcomes for this cohort, and further study of the genetics of early response is warranted.

Nonconcordance between genetic and clinical data suggests that definitions of refractoriness and remission should be more carefully and specifically defined, and take genetic correlates into consideration. Having said that, the statistical analyses were somewhat restricted, since only 10 patients qualified as treatment-refractory (Chiliza *et al.*, 2015a). Additionally, states of remission and refractoriness are notably difficult to define in clinical settings, since there is

extensive inter-individual heterogeneity in both schizophrenia symptom severity and antipsychotic response. For example, patients with high baseline PANSS scores can achieve a greater than 25% reduction with treatment, but still meet other criteria for nonresponse, since their endpoint PANSS scores will be high in relation to other individuals (Chiliza *et al.*, 2015a). Therefore, these loci should be investigated in a larger group of FES patients to i) provide insight into biological definitions of response states, and ii) investigate the potentially unique genetic mechanisms of treatment-refractoriness.

4.5.3.3 Metabolic outcomes

In comparison to PANSS outcomes, there were few associations with metabolic responses, i.e. changes in weight, BMI, HDL, LDL, total cholesterol, and triglycerides. Although none survived correction for multiple testing, the most significant association was between rs230539 and changes in total cholesterol ($P = 0.0096$). Furthermore, a 4q24 haplotype containing the major allele of this variant was significantly associated with changes in triglyceride levels ($P = 0.0237$; Table S5). Several other SNPs in this haplotype were nominally associated with metabolic changes. These results suggest a role for this haplotype in metabolic dysregulation, although this has not been previously observed. Another haplotype, A.T.C, containing the major alleles of rs10218843, rs11265461, and rs6427540, was also nominally associated with change in triglyceride levels ($P = 0.0492$). Similar to the 4q24 haplotype, this trio of SNPs also originates from the Liou *et al.* GWAS (2012), the first two minor alleles having been implicated in treatment-refractoriness and the third in LD with the pair. One possible link between refractoriness and increased metabolic ADRs is that patients who showed no response were consequently administered higher doses of flupenthixol; these increased doses may contribute more to adipogenic side effects than lower doses of the antipsychotic (Emsley, personal communication). Although the mechanisms are as yet undetermined, these regions should be further investigated for involvement in regulation of antipsychotic response and metabolic side effects.

The FES patients demonstrated considerable adverse metabolic changes after treatment. Indeed, AIWG and other metabolic side effects are common in antipsychotic-treated patients, and can have serious consequences such as metabolic syndrome and cardiovascular disease (Brennan, 2014; Chiliza *et al.*, 2015b). The lack of strong associations found within this study may be due to several factors. Firstly, only two rSNPs investigated were implicated in metabolic outcomes, namely rs6741819 (increase in triglycerides; Adkins *et al.*, 2011) and rs1352318 (increase in BMI; Athanasiu *et al.*, 2012). These two studies were not particularly well-characterised (refer to 3.5.1.1), making replication difficult. Secondly, while the PANSS is a standardised scale with high reliability

and validity (Levine *et al.*, 2011), several metabolic indices were measured, and techniques may have been inconsistent between measurements and/ or independent studies. Finally, this study focused on SNPs implicated in regulation. It may be the case that abnormal regulatory processes do not influence metabolic side effects to as great an extent as changes in coding regions do. Whatever the case, genome-wide studies accounting for coding and noncoding variation should be undertaken in order to identify genetic correlates and adequately manage metabolic-related ADRs.

4.5.4. GWAS comparisons

The variants originally associated with treatment outcomes in previous GWAS were included in the association analyses for two reasons: firstly, to directly compare outcomes in the GWAS cohorts with the FES cohort, and secondly, to assess the accuracy of a GWAS variant as a proxy for a region, by determining differences in association patterns between GWAS SNPs and corresponding LD SNPs. Several GWAS SNPs were significant for treatment response. For example, the haplotype on chromosome five contains two variants, rs17440909 and rs17742120, which were both previously associated with symptom severity as perceived by the patient (Clark *et al.*, 2013). Both SNPs, as well as the C.A and T.G haplotypes, were nominally associated with changes in PANSS Negative scores (Table S5). Opposite combinations of haplotypes reaching significance is unexpected. This finding suggests that the SNPs in the Clark *et al.* (2013) study served together as a tag for the causal region in their cohort. However, European ancestry individuals display much larger blocks of LD than Africans (Dalal *et al.*, 2010; Chimusa *et al.*, 2015), therefore this tag may be ineffective for the FES cohort, considering the ancestry make-up of the patients. This demonstrates the importance of population-specific studies in identifying causal variants.

Additionally, the SNP previously associated with treatment-refractoriness (rs230529; Liou *et al.*, 2012) in the large haplotype block on chromosome four (Figure 4.4), was not significant for any treatment outcomes in the FES cohort, despite having a relatively high frequency of 0.46. This haplotype and several of its individual variants were significantly associated with various response traits. Therefore, it is expected that the rs230529 SNP should reach significance in the FES cohort too, in order to function as an accurate proxy for the region. The r^2 values for this haplotype range from 0.22 to 0.96 (Figure 4.4), however when assessing D' confidence intervals, strong pairwise LD ($D' > 0.7$) is shown between all the SNPs within the haplotype (Figure S1). Differences between D' and r^2 values can be explained by their differing properties. D' is more sensitive to both allele frequency and sample size, and is usually inflated with a smaller cohort such as this one (Carlson *et al.*, 2004; Meadows *et al.*, 2008). There is debate over which measure should be used to define

LD between markers, but r^2 is generally preferred and accepted to be more applicable for association studies (Mueller, 2004; Zhao *et al.*, 2005).

The lack of associations for rs230529 within the South African cohort illustrates the importance of analysing each variant at a locus, since patterns of LD vary extensively between different population groups. This has implications for the way in which GWAS are currently designed and interpreted.

4.5.5. Study limitations

The greatest limitation of this study is the comparison of significant variants between different treatment cohorts. There are several factors that restrict direct comparison, and thus the results should be interpreted with caution and validated in other cohorts. Firstly, the differences in the types of antipsychotics the patients received has an impact on comparison of clinical outcomes. Although their equivalence in efficacy is still under debate, FGAs and SGAs produce vastly different side effect profiles, and ideally only drugs of the same class should be analysed together (Meltzer, 2013). Drug heterogeneity within previous GWAS, for example the studies assessing the CATIE cohort (Lieberman *et al.*, 2005), may also distort the clinical phenotype and prevent the identification of true associations. Furthermore, clinical assessments differ between studies, and clinical scores and classifications – even PANSS scores – are dependent on the evaluating psychiatrist to a certain extent. This has been illustrated by the lack of standard definitions for treatment refractoriness or remission (Chiliza *et al.*, 2015a).

Another limitation of this study is the small size of the FES cohort compared to the tens of thousands of individuals analysed by the PGC (Ripke *et al.*, 2014). However, it is important to note that the cohort is extremely well-characterised and homogenised. The patients were treatment-naïve at the commencement of the study, and all received the same antipsychotic via injection, ensuring adherence. It has been demonstrated that smaller clinical cohorts of well-characterised individuals have equivalent power to larger cohorts of less well-characterised patients (Samuels *et al.*, 2009). Furthermore, first episode cohorts provide increased power in pharmacogenetic studies (Zhang and Malhotra, 2013b). Considering these factors, this study minimised confounders and provided increased statistical power to detect associations (Reynolds, 2007), demonstrated by the emergence of significant findings that survived multiple testing, and the large effect size of 4.41%. The benefits of this cohort are particularly apparent when compared to the scarcity of similar cohorts available for the study of antipsychotic pharmacogenomics. The only exception is the study by Malhotra and

colleagues (2012), which demonstrated the ability to detect and replicate a genome-wide signal in a patient group of comparable size ($n = 139$) to the FES cohort.

4.6. **Conclusion**

This study successfully validated some of the bioinformatic results and demonstrated that rSNPs are important in antipsychotic response. There are several novel findings with regards to the potential mechanisms of treatment response in schizophrenia. Firstly, variation in the 4q24 region and *NFKB1* should be considered as novel targets when investigating antipsychotic mechanisms and heterogeneous treatment outcomes. Their connection to regulation, negative symptom severity, immunity, as well as the treatment-refractoriness endophenotype suggest that this locus has important and widespread implications in schizophrenia.

Additionally, this study demonstrated the importance of well-characterised cohorts and clear, standardised definitions of concepts such as remission, treatment-refractoriness, and early treatment response. The nonconcordance between genetic correlates and clinical classifications necessitates a re-evaluation of treatment outcome criteria. This is particularly important with regards to early response as a predictor of later positive outcomes. If accurate biomarkers can be uncovered, better first-line treatments can be applied and ADRs can be avoided. This, however, relies on extensive clinical characterisation.

Lastly, the Eurocentric nature of GWAS means that other population groups have been understudied. It is important to note that although not all bioinformatic candidates were found to be statistically significant in this study cohort, the absence of significance may just be representative of insufficient power to detect associations in this study, or that these loci may only be relevant in alternative population groups. The analysis of the FES cohort has contributed to narrowing the research gap between LMIC and developed countries, although the road ahead is long. Improvement in treatment outcomes calls for population-specific studies that benefit the individuals carrying the largest burden of disease.

CHAPTER 5: Conclusion and future perspectives

5.1. Conclusion

This study aimed to improve upon the interpretations of existing research on antipsychotic pharmacogenomics in schizophrenia by combining novel bioinformatic tools with validation within a previously underrepresented population group. The all-encompassing nature of this study – i.e. analysis of LD regions rather than only tag SNPs, inclusion of coding and noncoding variants, and evaluation of diverse, well-characterised clinical outcomes – has led to the formation of new hypotheses regarding the biology of antipsychotic mechanisms and treatment response. Firstly, many previously implicated SNPs were predicted to have roles in proximal, distal, and post-transcriptional regulation. Several have evidence for acting as eQTLs in the expression of genes previously not associated with treatment response. Perhaps the most important finding is that over half of the characterised rSNPs were predicted to affect the expression of *NFKB1*, which is itself a master regulator. This gene, as well as several other implicated proteins and motifs, have roles in immune functioning. Abnormal immune responses have long been suggested to contribute to the development of schizophrenia, and these results suggest that this may extend to the treatment outcomes of the disorder too. *NFKB1* is one of several examples of potential novel pharmacogenes. Interestingly, the region most significant for regulation was previously associated with treatment-refractoriness. The results of the bioinformatic analyses support the notion that treatment-refractoriness may be a unique endophenotype of schizophrenia, with distinct biological pathways leading to its development.

Importantly, the association analyses supported the hypothesised roles of regulatory regions in treatment response, as the majority of SNPs were significantly associated with at least one treatment outcome in the FES cohort. The weight of significant associations was found to be with the PANSS Negative outcome, which is an exciting avenue to follow since negative symptoms are the most chronic and pervasive, yet least understood of the symptom domains. The involvement of the 4q24 region in this respect suggests that there is a biological link between negative symptoms and treatment-refractoriness.

Interestingly, lending weight to the LD approach of the bioinformatic pipeline, there were instances in which the tag SNP and LD SNPs did not achieve the same or even similar patterns of associations within the cohort. This has important repercussions for the design and subsequent interpretation of GWAS studies. Furthermore, this study proved the importance of thorough clinical characterisation. Treatment response is complex and

heterogeneous, and a cohort must be well-characterised and homogenised as far as possible in order to arrive at robust conclusions regarding genetic correlates of drug response. The FES cohort is an example of such a cohort, particularly in comparison with previous GWAS cohorts. It is clear that GWAS patient recruitment and assessment must be improved so as not to dilute the phenotype. With this improvement, statistical power is increased and associations carry more weight.

The novel findings of this study suggest new directions for treatment response research in schizophrenia. Importantly, results of the association analyses suggest a new approach regarding characterisation of response states in schizophrenia. With the help of genetic studies, coupled with clinical guidelines such as RDoC (Insel *et al.*, 2010), classifying patient subgroups and response states based on biology can guide treatment strategies and improve long-term outcomes. Repeatedly following traditional hypotheses and candidate genes has been a popular approach, but has had limited success in underpinning the biology of drug response. The findings of this study, however, open new avenues for research by expanding focus to candidate gene pathways and networks, and have the potential to improve on a treatment approach that has been suboptimal for over 60 years.

5.2. Future perspectives

It is important to remember that GWAS is only one approach to uncovering the genetics of complex traits. GWAS are the hallmark for contributing knowledge to the CDCV hypothesis, but it is likely that a combination of rare, common, and *de novo* variation amalgamate to produce complex trait phenotypes such as antipsychotic response (van Dongen and Boomsma, 2013). In the future, a combination of studies, including large sample GWAS, whole genome sequencing, and subsequent functional validation should be employed. The key to uncovering variation of small effect size is a large study sample. Population-specific analyses of tens to hundreds of thousands of individuals, with replication, would advance pharmacogenomics as it has schizophrenia research (Ripke *et al.*, 2014).

Such large-scale studies generate massive amounts of data, often without including a post-analysis step of functional validation. This step is problematic, given the lack of a suitable *in vitro* model for schizophrenia or treatment response. An exciting area of research designed to address this is the engineering of induced pluripotent stem cell (iPSC)-derived neurons. With this approach, *in vitro* study of brain structures in different states is possible. For example, Brennand and colleagues (2011) reprogrammed fibroblasts from schizophrenia patients into neurons and found reduced neuronal connectivity and glutamate receptor

expression. Although only beginning, this type of research could allow for discovery of implicated pathways in action. The structural effects on neurons of regulatory variation can be investigated with this functional method and used to confirm results from predictive computational tools. Furthermore, a case/ control study in which iPSC-derived neurons are incubated with different antipsychotics could provide insight on the mechanisms and biological outcomes of FGAs and SGAs. Functional studies are essential for the discovery and implementation of accurate, clinically actionable biomarkers of treatment response.

Given the significance of thorough clinical characterisation for association studies, 18 European institutes recently joined forces to form the Optimization of Treatment and Management of Schizophrenia in Europe (OPTiMiSE) programme (<http://www.optimisetrialeu/>). This initiative aims to recruit 500 drug-naïve schizophrenia patients and perform a closely-monitored six-year trial of SGA treatment. Combined with genomic studies, the goal of this project is to optimise current treatment and develop new treatment strategies. This could provide unparalleled insight into early treatment response and antipsychotic mechanisms, and sets an example for other consortia by emphasising the importance of clinical stringency for genomic studies.

Closer to home, a promising new development in the study of previously underrepresented populations is the release of the African Genome Variation Project (Gurdasani *et al.*, 2015). The African LD populations included on publically available databases – such as YRI – are not necessarily accurate proxies for sub-Saharan African populations such as South Africans. Progress in research in LMIC is particularly important since these individuals are disproportionately burdened by communicable diseases and psychiatric disorders. Understudied and overburdened, research in this field would be immensely beneficial to individuals, communities, as well as societies and the economy of LMIC. Pharmacogenomics has the power to achieve some of these goals. The results of Gurdasani and colleagues' (2015) research add more variation to the growing pool of genetic information on understudied groups, and allow for improvement upon the novel bioinformatic pipeline. What is more, this pipeline needs not be restricted to antipsychotic treatment response; it is applicable to any complex genetic disorder, and is an agnostic approach with the potential to reveal novel dysregulated pathways and ultimately improve disease outcomes.

REFERENCES

- The 1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. *Nature*, 468, 1061-73.
- Åberg, K., Adkins, D.E., Bukszár, Webb, B.T., Caroff, S.N., Miller, D.D., *et al.* 2010. Genomewide association study of movement-related adverse antipsychotic effects. *Biological Psychiatry*, 67(3), 279-82.
- Åberg, K., Adkins, D.E., Liu, Y., McClay, J.L., Bukszár, J., Jia, P. 2012. Genome-wide association study of antipsychotic induced QTc interval prolongation. *Pharmacogenomics*, 12(2), 165-72.
- Adam, D. 2013. On the spectrum. *Nature*, 496(7446), 416-18.
- Adeyemo, A., Rotimi, C. 2010. Genetic variants associated with complex human diseases show wide variation across multiple populations. *Public Health Genomics*, 13(2), 72-9.
- Adkins, D.E., Åberg, K., McClay, J.L., Bukszár, J., Zhao, Z., Jia, P., *et al.* 2011. Genomewide pharmacogenomic study of metabolic side effects to antipsychotic drugs. *Molecular Psychiatry*, 16(3), 321-2.
- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., *et al.* 2010. A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4), 248-9.
- Agid, O., Kapur, S., Arenovich, T., Zipursky, R.B. 2003. Delayed-onset hypothesis of antipsychotic action: a hypothesis tested and rejected. *Archives of General Psychiatry*, 60(12), 1228-35.
- Albus, M. 2012. Clinical courses of schizophrenia. *Pharmacopsychiatry*, 45(S1), S31-5.
- Alexander, D.H., Novembre, J., Lange, K. 2010. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655-64.
- Alkelai, A., Greenbaum, L., Rigbi, A., Kanyas, K., Lerer, B. 2009. Genome-wide association study of antipsychotic-induced parkinsonism severity among schizophrenia patients. *Psychopharmacology*, 206(3), 491-9.
- Allen, N.C., Bagade, S., McQueen, M.B., Ioannidis, J.P.A., Kavvoura, F.K., Khoury, M.J., *et al.* 2008. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nature Genetics*, 40(7), 827-34.
- Alvarez-Rodriguez, J., Alvarez-Silva, S., Alvarez-Silva, I. 2014. Is the current diagnosis of schizophrenia useful or harmful? *Open Journal of Medical Psychology*, 3, 157-60.
- Alvir, J.M., Lieberman, J.A., Safferman, A.Z., Schwimmer, J.L., Schaaf, J.A. 1993. Clozapine-induced agranulocytosis. Incidence and risk factors in the United States. *New England Journal of Medicine*, 329(3), 162-7.

- American Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. 1st ed. Washington, DC: American Psychiatric Association.
- American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. 5th ed. Washington, DC: American Psychiatric Association.
- Andreasen, N.C. 1983. Scale for the assessment of negative symptoms (SANS). The University of Iowa, Iowa City.
- Andreasen, N.C. 1984. Scale for the assessment of positive symptoms (SAPS). The University of Iowa, Iowa City.
- Andreasen, N.C., Carpenter, W.T. Jr., Kane, J.M., Lasser, R.A., Marder, S.R., Weinberger, D.R. 2005. Remission in schizophrenia: proposed criteria and rationale for consensus. *American Journal of Psychiatry*, 162, 441–9.
- Arnedo, J., Svrakic, D.M., del Val, C., Romero-Zaliz, R., Hernández-Cuervo, H., Molecular Genetics of Schizophrenia Consortium, *et al.* 2014. Uncovering the hidden risk architecture of the schizophrenias: confirmation of three independent genome-wide association studies. *American Journal of Psychiatry*, 172(2), 139-53.
- Arranz, M.J., de Leon, J. 2007. Pharmacogenetics and pharmacogenomics of schizophrenia: a review of last decade of research. *Molecular Psychiatry*, 12, 707-47.
- Arranz, M.J., Munro, J.C. 2011. Toward understanding genetic risk for differential antipsychotic response in individuals with schizophrenia. *Expert Reviews on Clinical Pharmacology*, 4(3), 389-405.
- Ascher-Svanum, H., Zhu, B., Faries, D.E., Salkever, D., Slade, E.P., Peng, X., Conley, R.R. 2010. The cost of relapse and the predictors of relapse in the treatment of schizophrenia. *BMC Psychiatry*, 10:2, 1-7.
- Athanasou, L., Brown, A.A., Birkenaes, A.B., Mattingsdal, M., Agartz, I., Melle, I., *et al.* 2012. Genome-wide association study identifies genetic loci associated with body mass index and high density lipoprotein-cholesterol levels during psychopharmacological treatment. *Psychiatry Research*, 197(3), 327-36.
- Auerbach, R.K., Chen, B., Butte, A.J. 2013. Relating genes to function: identifying enriched transcription factors using the ENCODE ChIP-Seq significance tool. *Bioinformatics*, 29(15), 1922-4.
- Bank, S., Andersen, P.S., Burisch, J., Pedersen, N., Roug, S., Galsgaard, J., *et al.* 2014. Associations between functional polymorphisms in the NFκB signaling pathway and response to anti-TNF treatment in Danish patients with inflammatory bowel disease. *The Pharmacogenomics Journal*, 14(6), 526-34.
- Barch, D.M., Bustillo, J., Gaebel, W., Gur, R., Heckers, S., Malaspina, D., *et al.* 2013. Logic and justification for dimensional assessment of symptoms and related clinical

- phenomena in psychosis: relevance to DSM-5. *Schizophrenia Research*, 150(1), 15-20.
- Barrett, J.C., Fry, B., Maller, J., Daly, M.J. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), 263–5.
- Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V., Izaurralde, E. 2007. A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *The EMBO Journal*, 26(6), 1591–601.
- Bertram, L. 2008. Genetic research in schizophrenia: new tools and future perspectives. *Schizophrenia Bulletin*, 34(5), 806-12.
- Blanc, O., Brousse, G., Meary, A., Leboyer, M., Llorca, P-M. 2010. Pharmacogenetics of response efficacy to antipsychotics in schizophrenia: pharmacodynamic aspects. Review and implications for clinical research. *Fundamental & Clinical Pharmacology*, 24, 139-160.
- Bloom, D.E., Cafiero, E.T., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L.R., Fathima, S., *et al.* 2011. *The Global Economic Burden of Noncommunicable Diseases*. Geneva: World Economic Forum.
- Bonham, C., Abbott, C., 2008. Are second-generation antipsychotics a distinct class? *Journal of Psychiatric Practice*, 14, 225–37.
- Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., *et al.* 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research*, 22(9), 1790-7.
- Brandl, E.J., Kennedy, J.L., Müller, D.J. 2014. Pharmacogenetics of antipsychotics. *Canadian Journal of Psychiatry*, 59(2), 76-88.
- Brennan, M. 2014. Pharmacogenetics of second-generation antipsychotics. *Pharmacogenomics*, 15(6), 869-84.
- Brennand, K.J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., *et al.* 2011. Modelling schizophrenia using human induced pluripotent stem cells. *Nature*, 473(7346), 221-5.
- Cacabelos, R., Hashimoto, R., Takeda, M. 2011. Pharmacogenomics of antipsychotics efficacy for schizophrenia. *Psychiatry and Clinical Neurosciences*, 65(1), 3-19.
- Campbell, M.C., Tishkoff, S.A. 2008. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annual Review of Genomics and Human Genetics*, 9, 403-33.
- Cantor-Graae, E., Selten, J.P. 2005. Schizophrenia and migration: a meta-analysis and review. *American Journal of Psychiatry*, 162, 12–24.
- Carlson, C.S., Eberle, M.A., Rieder, M.J., Yi, Q., Kruglyak, L., Nickerson, D.A. 2004. Selecting a maximally informative set of single-nucleotide polymorphisms for

- association analyses using linkage disequilibrium. *American Journal of Human Genetics*, 74, 106-20.
- Carlsson, A., Lindqvist, M., 1963. Effect of chlorpromazine or haloperidol on formation of 3-methoxytyramine and normetanephrine in mouse brain. *Acta Pharmacologica et Toxicologica*, 20 140–4.
- Carpenter, W.T. Jr., Davis, J.M. 2012. Another view of the history of antipsychotic drug discovery and development. *Molecular Psychiatry*, 17(12), 1168-73.
- Carr, D.F., Alfirevic, A., Pirmohamed, M. 2014. Pharmacogenomics: current state-of-the-art. *Genes*, 5(2), 430-43.
- Case, M., Stauffer, V.L., Ascher-Svanum, H., Conley, R., Kapur, S., Kane, J.M., *et al.* 2011. The heterogeneity of antipsychotic response in the treatment of schizophrenia. *Psychological Medicine*, 41(6), 1291-1300.
- Caspi, A., Davidson, M., Tamminga, C.A. 2004. Treatment-refractory schizophrenia. *Dialogues in Clinical Neuroscience*, 6(1), 61-70.
- Chiliza, B., Asmal, L., Kilian, S., Phahladira, L., Emsley, R. 2015a. Rate and predictors of non-response to first-line antipsychotic treatment in first-episode schizophrenia. *Human Psychopharmacology: Clinical & Experimental*, 30(3), 173-82.
- Chiliza, B., Asmal, L., Oosthuizen, P., van Niekerk, E., Erasmus, R., Kidd, M., *et al.* 2015b. Changes in body mass and metabolic profiles in patients with first-episode schizophrenia treated for 12 months with a first-generation antipsychotic. *European Psychiatry*, 30(2), 277-83.
- Chimusa, E.R., Zaitlen, N., Daya, M., Möller, M., van Helden, P.D., Mulder, N.J., *et al.* 2014. Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. *Human Molecular Genetics*, 23(3), 796-809.
- Chimusa, E.R., Meintjies, A., Tchanga, M., Mulder, N., Seioqhe, C., Soodyall, H., Ramesar, R. 2015. A genomic portrait of haplotype diversity and signatures of selection in indigenous southern African populations. *PLoS Genetics*, 11(3), e1005052.
- Christoforou, A., Dondrup, M., Mattingsdal, M., Mattheisen, M., Giddaluru, S., Nöthen, M.M., *et al.* 2012. Linkage-disequilibrium-based binning affects the interpretation of GWASs. *American Journal of Human Genetics*, 90(4), 727-33.
- Chowdhury, N., Remington, G., Kennedy, J. 2011. Genetics of antipsychotic-induced side effects and agranulocytosis. *Current Psychiatry Reports*, 13(2), 156-65.
- Chung, C.C., Kanetsky, P.A., Wang, Z., Hildebrandt, M.A.T., Koster, R., Skotheim, R.I., *et al.* 2013. Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nature Genetics*, 45(6), 680-5.
- Clark, S.L., Souza, R.P., Adkins, D.E., Åberg, K., Bukszar, J., McClay, J.L., *et al.* 2013. Genome-wide association study of patient-rated and clinician-rated global impression

- of severity during antipsychotic treatment. *Pharmacogenetics and Genomics*, 23(2), 69-77.
- Clarke, M.C., Kelleher, I., Clancy, M., Cannon, M. 2012. Predicting risk and the emergence of schizophrenia. *Psychiatric Clinics of North America*, 35(3), 585-612.
- Collins, P.Y., Patel, V., Joestl, S.S., March, D., Insel, T.R., Daar, A.S. 2011. Grand challenges in global mental health. *Nature*, 475(7354), 27-30.
- Collins, A.L., Kim, Y., Sklar, P., International Schizophrenia Consortium, O'Donovan, M.C., Sullivan, P.F. 2012. Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results. *Psychological Medicine*, 42(3), 607-16.
- Conley, R.R., Kelly, D.L. 2001. Management of treatment resistance in schizophrenia. *Biological Psychiatry*, 50(11), 898-911.
- Cook, J., Hunter, G., Vernon, J. 2009. The future costs, risks, and rewards of drug development. *Pharmacoeconomics*, 27, 355-63.
- Cookson, W., Liang, L., Abecasis, G., Moffatt, M., Lathrop, M. 2009. Mapping complex disease traits with global gene expression. *Nature Reviews Genetics*, 10(3), 184-94.
- Cooper, G., Shendure, J. 2011. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature Reviews Genetics*, 12(9), 628-40.
- Correll, C.U., Lencz, T., Malhotra, A.K. 2011. Antipsychotic drugs and obesity. *Trends in Molecular Medicine*, 17(2), 97-107.
- Curtis, D. 2013. Consideration of plausible genetic architectures for schizophrenia and implications for analytic approaches in the era of next generation sequencing. *Psychiatric Genetics*, 23(1), 1-10.
- Dalal, S., Holmes, M.D., Ramesar, R.S. 2010. Advancing public health genomics in Africa through prospective cohort studies. *Journal of Epidemiology and Community Health*, 64(7), 585-6.
- Daya, M., van der Merwe, L., Galal, U., Möller, M., Salie, M., Chimusa, E.R., *et al.* 2013. A panel of ancestry informative markers for the complex five-way admixed South African coloured population. *PloS One*, 8(12), e82224.
- de Jesus Mari, J., Patel, V., Kieling, C., Anders, M., Jakovljevi, M., Lam, L.C.W., *et al.* 2009. The 5/95 Gap on the dissemination of mental health research: The World Psychiatric Association (WPA) task force report on project with editors of low and middle income (LAMI) countries. *African Journal of Psychiatry*, 12(1), 33-9.
- de Leon, J. 2009. The future (or lack of future) of personalized prescription in psychiatry. *Pharmacological Research*, 59(2), 81-9.
- Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., Gasquet, I., Kovess, V., Lepine, J.P., *et al.* 2004. Prevalence, severity and unmet needs for treatment of mental disorders in

- the World Health Organization World Mental Health Surveys. *Journal of the American Medical Association*, 291(21), 2581-90.
- de Wit, E., Delpont, W., Rugamika, C.E., Meintjes, A., Möller, M., van Helden, P., *et al.* 2010. Genome-wide analysis of the structure of the South African Coloured population in the Western Cape. *Human Genetics*, 128(2), 145-53.
- Doherty, J.L., Owen, M.J. 2014. Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. *Genomic Medicine*, 6(4), 29-41.
- Dorado, P., Penas-Lledo, E.M., de la Rubia, A. 2009. Relevance of CYP2D6 -1584C>G polymorphism for thioridazine:mesoridazine plasma concentration ratio in psychiatric patients. *Pharmacogenomics*, 10(7), 1083–9.
- Drögemöller, B.I., Wright, G.E.B., Niehaus, D.J.H., Emsley, R.A., Warnich, L. 2011. Whole-genome resequencing in pharmacogenomics: moving away from past disparities to globally representative applications. *Pharmacogenomics*, 12(12), 1717-28.
- Drögemöller, B.I. 2013. Investigation of genetic variation contributing to antipsychotic treatment response in a South African first episode schizophrenia cohort. Unpublished doctoral dissertation. Stellenbosch: Stellenbosch University. Available: <http://hdl.handle.net/10019.1/95473>.
- Drögemöller, B.I., Wright, G.E.B., Warnich, L. 2014a. Considerations for rare variants in drug metabolism genes and the clinical implications. *Expert Opinion on Drug Metabolism & Toxicology*, 10(6), 873-84.
- Drögemöller, B.I., Niehaus, D.J.H., Chiliza, B., van der Merwe, L., Asmal, L., Malhotra, A.K., *et al.* 2014b. Patterns of variation influencing antipsychotic treatment outcomes in South African first-episode schizophrenia patients. *Pharmacogenomics*, 15(2), 189-99.
- Ecker, J.R., Bickmore, W.A., Barroso, I., Pritchard, J.K., Gilad, Y., Segai, E. 2012. Genomics: ENCODE explained. *Nature*, 489(7414), 52-5.
- Emsley, R., Chiliza, B., Asmal, L., Lehloeny, K. 2011. The concepts of remission and recovery in schizophrenia. *Current Opinion in Psychiatry*, 24(2), 114-21.
- Emsley, R., Oosthuizen, P., Koen, L., Niehaus, D., Martinez, L. 2013. Comparison of treatment response in second-episode versus first-episode schizophrenia. *Journal of Clinical Psychopharmacology*, 33(1), 80-3.
- The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447, 799–816.
- The ENCODE Project Consortium. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biology*, 9(4). Available: <http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1001046#pbio-1001046-g008>.

- The FANTOM Consortium, The RIKEN PMI, CLST (DGT). 2014. A promoter-level mammalian expression atlas. *Nature*, 507(7493), 462-70.
- Flint, J., Munafò, M. 2014. Schizophrenia: genesis of a complex disease. *Nature*, 511, 412-13.
- Flockhart, D.A., O’Kane, D., Williams, M.S., Watson, M.S., Gage, B., Gandolfi, R., *et al.* 2008. Pharmacogenetic testing of *CYP2C9* and *VKORC1* alleles for warfarin. *Genetics in Medicine*, 10(2), 139–50.
- Fond, G., d’Albis, M.-A., Jamain, S., Tamouza, R., Arango, C., Wolfgang Fleischhacker, W., *et al.* 2015. The promise of biological markers for treatment response in first-episode psychosis: a systematic review. *Schizophrenia Bulletin*, 41(3), 559-73.
- Freedman, M.L., Monteiro, A.N.A., Gayther, S.A., Coetzee, G.A., Risch, A., Plass, C., *et al.* 2011. Principles for the post-GWAS functional characterization of cancer risk loci. *Nature Genetics*, 43(6), 513-8.
- Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., *et al.* 2002. The structure of haplotype blocks in the human genome. *Science*, 296(5576), 2225-9.
- Gaedigk, A., Coetsee, C. 2008. The *CYP2D6* gene locus in South African Coloureds: unique allele distributions, novel alleles and gene arrangements. *European Journal of Clinical Pharmacology*, 64, 465-75.
- Georgitsi, M., Zukic, B., Pavlovic, S., Patrinos, G.P. 2011. Transcriptional regulation and pharmacogenomics. *Pharmacogenomics*, 12(5), 655-73.
- Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., *et al.* 2011. Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nature Genetics*, 43, 860–3.
- Girard, S.L., Dion, P.A., Rouleau, G.A. 2012. Schizophrenia genetics: putting all the pieces together. *Current Neurology and Neuroscience Reports*, 12(3), 261-6.
- Global Forum for Health Research. 2000. *10/90 report on health research 2000*. Geneva: Global Forum for Health Research. Available: <http://www.globalforumhealth.org/about/1090-gap/>.
- Graur, D., Zheng, Y., Price, N., Azevedo, R.B.R., Zufall, R.A., Elhaik, E. 2013. On the immortality of television sets: “function” in the human genome according to the evolution-free gospel of ENCODE. *Genome Biology and Evolution*, 5(3), 578-90.
- Guo, L., Du, Y., Chang, S., Zhang, K., Wang, J. 2014. rSNPBase: a database for curated regulatory SNPs. *Nucleic Acids Research*, 42. Available: <http://nar.oxfordjournals.org/content/early/2013/11/26/nar.gkt1167.full>.
- Gupta, S., Venkatesh, A., Ray, S., Srivastava, S. 2014. Challenges and prospects for biomarker research: a current perspective from the developing world. *Biochimica Et Biophysica Acta*, 1844(5), 899-908.

- Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., *et al.* 2015. The African Genome Variation Project shapes medical genetics in Africa. *Nature*, 517(7534), 327-32.
- Halene, T.B., Siegel, S.J. 2008. Antipsychotic-like properties of phosphodiesterase 4 inhibitors: evaluation of 4-(3-butoxy-4-methoxybenzyl)-2-imidazolidinone (RO-20-1724) with auditory event-related potentials and prepulse inhibition of startle. *Journal of Pharmacology and Experimental Therapeutics*, 326, 230-9.
- Hall, G., Singh, I.S., Hester, L., Hasday, J.D., Rogers, T.B. 2005. Inhibitor-kappaB kinase-beta regulates LPS-induced TNF-alpha production in cardiac myocytes through modulation of NF-kappaB p65 subunit phosphorylation. *American Journal of Physiology Heart and Circulatory Physiology*, 289, 2103-11.
- Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M. Mehta, J.P., Collins, F.S., Manolio, T.A. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, 106, 9362-7.
- Hinds, D.A., Stuve, L.L., Nilsen, G.B., Halperin, E., Eskin, E., Ballinger, D.G., *et al.* 2005. Whole-genome patterns of common DNA variation in three human populations. *Science*, 307, 1072-9.
- Huang, D.W., Sherman, B.T., Lempicki, R.A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44-57.
- Huynh, T., Khan, J.M., Ranganathan, S. 2011. A comparative structural bioinformatics analysis of inherited mutations in β -D-mannosidase across multiple species reveals a genotype-phenotype correlation. *BMC Genomics*, 12(S3), 1-13.
- Ikediobi, O., Aouizerat, B., Xiao, Y., Gandhi, M., Gebhardt, S., Warnich, L. 2011. Analysis of pharmacogenetic traits in two distinct South African populations. *Human Genomics*, 5, 265-82.
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D., Quinn, K., *et al.* 2010. Research Domain Criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 167(7), 748-51.
- Insel, T.R., Cuthbert, B.N. 2015. Brain disorders? Precisely. *Science*, 348(6234), 499-500.
- The International HapMap Consortium. 2003. The international HapMap project. *Nature*, 426, 789-96.
- International Warfarin Pharmacogenetics Consortium. 2009. Estimation of the warfarin dose with clinical and pharmacogenetic data. *New England Journal of Medicine*, 360(8), 753-64.

- Iyegbe, C., Campbell, D., Butler, A., Ajnakina, O., Sham, P. 2014. The emerging molecular architecture of schizophrenia, polygenic risk scores and the clinical implications for GxE research. *Social Psychiatry & Psychiatric Epidemiology*, 49(2), 169-82.
- Jeppesen, P., Petersen, L., Thorup, A., Abel, M-B., Øhlenschläger, J., Christensen, T. Ø., *et al.* 2008. The association between pre-morbid adjustment, duration of untreated psychosis and outcome in first-episode psychosis. *Psychological Medicine*, 38(8), 1157-66.
- Jones, P.B., Barnes, T.R.E., Davies, L., Dunn, G., Lloyd, H., Hayhurst, K.P., *et al.* 2006. Randomized controlled trial of the effect on quality of life of second- vs first-generation antipsychotic drugs in schizophrenia: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1). *Archives of General Psychiatry*, 63, 1079- 87.
- Jungerius, B.J., Hoogendoorn, M.L., Bakker, S.C., Van't Slot, R., Bardoel, A.F., Ophoff, R.A., *et al.* 2007. An association screen of myelin-related genes implicates the chromosome 22q11 *PIK4CA* gene in schizophrenia. *Molecular Psychiatry*, 13(11), 1060-8.
- Juraeva, D., Haenisch, B., Zapatka, M., Frank, J., GROUP Investigators, iPSYCH-GEMS SCZ working group, *et al.* 2014. Integrated pathway-based approach identifies association between genomic regions at *CTCF* and *CACNB2* and schizophrenia. *PLoS Genetics*, 10(6), e1004345.
- Kadri, N., Sartorius, N. 2005. The global fight against the stigma of schizophrenia. *PLoS Medicine*, 2(7). Available: <http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.0020136>.
- Kahn, R.S., Sommer, I.E. 2015. The neurobiology and treatment of first-episode schizophrenia. *Molecular Psychiatry*, 20, 84-97.
- Kameyama, Y., Yamashita, K., Kobayashi, K., Hosokawa, M., Chiba, K. 2005. Functional characterization of *SLCO1B1* (OATP-C) variants, *SLCO1B1*5*, *SLCO1B1*15* and *SLCO1B1*15+C1007G*, by using transient expression systems of HeLa and HEK293 cells. *Pharmacogenetics and Genomics*, 15(7), 513–22.
- Kapur, S., Mamo, D. 2003. Half a century of antipsychotics and still a central role for dopamine D₂ receptors. *Progress In Neuro-Psychopharmacology & Biological Psychiatry*, 27(7), 1081-90.
- Kay, S.R., Fiszbein, A., Opler, L.A. 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, 13(2), 261-76.
- Kerem, B., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., *et al.* 1989. Identification of the cystic fibrosis gene: genetic analysis. *Science*, 245(4922), 1073-80.

- Kim, Y., Zerwas, S., Trace, S.E., Sullivan, P.F. 2011. Schizophrenia genetics: where next? *Schizophrenia Bulletin*, 37(3), 456-63.
- Kim, Y.S., State, M.W. 2014. Recent challenges to the psychiatric diagnostic nosology: a focus on the genetics and genomics of neurodevelopmental disorders. *International Journal of Epidemiology*, 43(2), 465-75.
- Kingsmore, S.F., Lindquist, I.E., Mudge, J., Gessler, D.D., Beavis, W.D. 2008. Genome-wide association studies: progress and potential for drug discovery and development. *Nature Reviews Drug Discovery*, 7(3), 221-30.
- Kinon, B.J., Chen, L., Ascher-Svanum, H., Stauffer, V.L., Kollack-Walker, S., Sniadecki, J.L., et al. 2008. Predicting response to atypical antipsychotics based on early response in the treatment of schizophrenia. *Schizophrenia Research*, 102, 230–40.
- Kinon, B.J., Chen, L., Ascher-Svanum, H., Stauffer, V.L., Kollack-Walker, S., Zhou, W., et al. 2010. Early response to antipsychotic drug therapy as a clinical marker of subsequent response in the treatment of schizophrenia. *Neuropsychopharmacology*, 35, 581–90.
- Kirkpatrick, B., Fenton, W.S., Carpenter, W.T., Marder, S.R., 2006. The NIMH- MATRICS consensus statement on negative symptoms. *Schizophrenia Bulletin*, 32, 214–19.
- Klein, K., Zanger, U. 2013. Pharmacogenomics of cytochrome P450 3A4: recent progress toward the “missing heritability” problem. *Frontiers in Genetics*, 4(12), 1-15.
- Kodzius, R., Kojima, K., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., et al. 2006. CAGE: cap analysis of gene expression. *Nature Methods*, 3(3), 211-22.
- Krabbendam, L., van Os, J. 2005. Schizophrenia and urbanicity: a major environmental influence - conditional on genetic risk. *Schizophrenia Bulletin*, 31, 795–9.
- Kumar, P., Henikoff, S., Ng, P.C. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(7), 1073-81.
- Kuroiwa, M., Snyder, G.L., Shuto, T., Fukuda, A., Yanagawa, Y., Benavides, D.R., et al. 2011. Phosphodiesterase 4 inhibition enhances the dopamine D1 receptor/ PKA/DARPP-32 signaling cascade in frontal cortex. *Psychopharmacology*, 219, 1065-79.
- Lavedan, C., Licamele, L., Volpi, S., Hamilton, J., Heaton, C., Mack, K., et al. 2009. Association of the NPAS3 gene and five other loci with response to the antipsychotic iloperidone identified in a whole genome association study. *Molecular Psychiatry*, 14(8), 804-19.
- Lebedeva, S. 2012. Transcriptome-wide functional analysis of post-transcriptional regulatory interactions of the RNA-binding protein HuR/ELAVL1. Unpublished doctoral dissertation. Berlin: Freie Universität Berlin. Available: http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000038642.

- León-Cachón, R.B.R., Ascacio-Martínez, J.A., Barrera-Saldaña, H.A. 2012. Individual response to drug therapy: bases and study approaches. *Revista de Investigación Clínica*, 64(4), 364-76.
- Leucht, S., Heres, S., Hamann, J., Kane, J.M. 2008. Methodological issues in current antipsychotic drug trials. *Schizophrenia Bulletin*, 34, 275–85.
- Leucht, S., Corves, C., Arbter, D., Engel, R.R., Li, C., Davis, J.M. 2009. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. *The Lancet*, 373, 31–41.
- Levine, S.Z., Rabinowitz, J., Rizopoulos, D. 2011. Recommendations to improve the Positive and Negative Syndrome Scale (PANSS) based on item response theory. *Psychiatry Research*, 188(3), 446-52.
- Lewis, S., Lieberman, J. 2008. CATIE and CUtLASS: can we handle the truth? *British Journal of Psychiatry*, 192(3), 161-3.
- Li-Wan-Po, A., Girard, T., Farndon, P., Cooley, C., Lithgow, J. 2010. Pharmacogenetics of CYP2C19: functional and clinical implications of a new variant CYP2C19*17. *British Journal of Clinical Pharmacology*, 69(3), 222–30.
- Lieberman, J.A., Perkins, D., Belger, A., Chakos, M., Jaskog, F., Boteva, K., Gilmore, J. 2001. The early stages of schizophrenia: speculations on pathogenesis, pathophysiology, and therapeutic approaches. *Biological Psychiatry*, 50, 884-97.
- Lieberman, J.A., Stroup, S., McEvoy, J., Swartz, M.S., Rosenheck, R.A., Perkins, D.O., *et al.* 2005. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine*, 353, 1209-23.
- Lin, C-H., Huang, C-L., Chang, Y-C., Chen, P-W., Lin, C-Y., Tsai, G.E., Lane, H-Y. 2013. Clinical symptoms, mainly negative symptoms, mediate the influence of neurocognition and social cognition on functional outcome of schizophrenia. *Schizophrenia Research*, 146, 231-7.
- Liou, Y-J., Wang, H-H., Lee, M-T.M., Wang, S-C., Chiang, H-L., Chen, C-C., *et al.* 2012. Genome-wide association study of treatment refractory schizophrenia in Han Chinese. *PloS One*, 7(3), e33598.
- Lohoff, F.W., Ferraro, T.N. 2010. Pharmacogenetic considerations in the treatment of psychiatric disorders. *Expert Opinion on Pharmacotherapy*, 11(3), 423-39.
- Lopez-Munoz, F., Alamo, C., Cuenca, E., Shen, W.W., Clervoy, P., Rubio, G. 2005. History of the discovery and clinical introduction of chlorpromazine. *Annals of Clinical Psychiatry*, 17, 113–35.
- Lund, C., Kleintjes, S., Kakuma, R., Flisher, A.J. 2010. Public sector mental health systems in South Africa: inter-provincial comparisons and policy implications. *Social Psychiatry and Psychiatric Epidemiology*, 45(3), 393-404.

- Mah, S., Nelson, M.R., Delisi, L.E., Reneland, R.H., Markward, N., James, M.R., *et al.* 2006. Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Molecular Psychiatry*, 11, 471-78.
- Malaspina, D., Walsh-Messinger, J., Gaebel, W., Smith, L.M., Gorun, A., Prudent, V., *et al.* 2014. Negative symptoms, past and present: a historical perspective and moving to DSM-5. *European Neuropsychopharmacology*, 24(5), 710-24.
- Malhotra, A.K., Correll, C.U., Chowdhury, N.I., Müller, D.J., Gregersen, P.K., Lee, A.T., *et al.* 2012. Association between common variants near the melanocortin 4 receptor gene and severe antipsychotic drug-induced weight gain. *Archives of General Psychiatry*, 69(9), 904-12.
- Mamdani, F., Martin, M.V., Lencz, T., Rollins, B., Robinson, D.G., Moon, E.A., *et al.* 2013. Coding and noncoding gene expression biomarkers in mood disorders and schizophrenia. *Disease Markers*, 35(1), 11-21.
- Manke, T., Heinig, M., Vingron, M. 2010. Quantifying the effect of sequence variation on regulatory interactions. *Human Mutation*, 31(4), 477-83.
- Marigo, I., Bosio, E., Solito, S., Mesa, C., Fernandez, A., Dolcetti, L., *et al.* 2010. Tumor-induced tolerance and immune suppression depend on the C/EBP β transcription factor. *Immunity*, 32(6), 790-802.
- Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., *et al.* 2013. Systematic localization of common disease-associated variation in regulatory DNA. *Science*, 337, 1190-5.
- McCarthy, S.E., McCombie, W.R., Corvin, A. 2014. Unlocking the treasure trove: from genes to schizophrenia biology. *Schizophrenia Bulletin*, 40(3), 492-6.
- McClay, J.L., Adkins, D.E., Åberg, K., Bukszár, J., Khachane, A.N., Keefe, R.S.E., *et al.* 2011a. Genome-wide pharmacogenomics study of neurocognition as an indicator of antipsychotic treatment response in schizophrenia. *Neuropsychopharmacology*, 36(3), 616-26.
- McClay, J.L., Adkins, D.E., Åberg, K., Stroup, S., Perkins, D.O., Vladimirov, V.I., *et al.* 2011b. Genome-wide pharmacogenomic analysis of response to treatment with antipsychotics. *Molecular Psychiatry*, 16, 76-85.
- McEvoy, J.P., Hogarty, G.E., Steingard, S. 1991. Optimal dose of neuroleptic in acute schizophrenia. A controlled study of the neuroleptic threshold and higher haloperidol use. *Archives of General Psychiatry*, 48(8), 739-45.
- McEvoy, J.P., Meyer, J.M., Goff, D.C., Nasrallah, H.A., Davis, S.M., Sullivan, L., *et al.* 2005. Prevalence of the metabolic syndrome in patients with schizophrenia: baseline results from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE)

- schizophrenia trial and comparison with national estimates from NHANES III. *Schizophrenia Research*, 80(1), 19-32.
- McEvoy, J.P., Lieberman, J.A., Stroup, T.S., Davis, S.M., Meltzer, H.Y., Rosenheck, R.A., *et al.* 2006. Effectiveness of clozapine versus olanzapine, quetiapine, and risperidone in patients with chronic schizophrenia who did not respond to prior antipsychotic treatment. *American Journal of Psychiatry*, 163, 600–10.
- McKusick, V.A. 1998. *Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders*, 12th ed. Johns Hopkins University Press, Baltimore, MD.
- Meadows, J.R.S., Chan, E.K.F., Kijas, J.W. 2008. Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genetics*, 9, 61.
- Meltzer, H.Y. 2013. Update on typical and atypical antipsychotics drugs. *Annual Review of Medicine*, 64, 393-406.
- Millan, M.J., Fone, K., Steckler, T., Horan, W.P. 2014. Negative Symptoms of schizophrenia: clinical characteristics, pathophysiological substrates, experimental models and prospects for improved treatment. *European Neuropsychopharmacology*, 24(5), 645-92.
- Miller, S.A., Dykes, D.D., Polesky, H.F. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, 16(3), 1215.
- Mitchell, C., Gregerson, N., Krause, A. 2011. Novel *CYP2C9* and *VKORC1* gene variants associated with warfarin dosage variability in the South African black population. *Pharmacogenomics*, 12(7), 953-63.
- Mowry, B., Gratten, J. 2013. The emerging spectrum of allelic variation in schizophrenia: current evidence and strategies for the identification and functional characterization of common and rare variants. *Molecular Psychiatry*, 18(1), 38-52.
- Mueller, J.C. 2004. Linkage disequilibrium for different scales and applications. *Briefings in Bioinformatics*, 5(4), 355-64.
- Mueser, K.T., McGurk, S.R. 2004. Schizophrenia. *The Lancet*, 363(9426), 2063-72.
- Muir, W.J., Gosden, C.M., Brookes, A.J., Fantes, J., Evans, K.L., Maguire, S.M., *et al.* 1995. Direct microdissection and microcloning of a translocation breakpoint region, t(1;11)(q42.2;q21), associated with schizophrenia. *Cytogenetics and Cell Genetics*, 70, 35–40.
- Murali, V., Oyeboode, F. 2004. Poverty, social inequality and mental health. *Advances in Psychiatric Treatment*, 10(3), 216-24.
- Murray, C.J.L., Lopez, A.D. 1996. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Cambridge, Massachusetts.

- Murayama, N., Soyama, A., Saito, Y., Nakajima, Y., Komamura, K., Ueno, K., *et al.* 2004. Six novel nonsynonymous *CYP1A2* gene polymorphisms: catalytic activities of the naturally occurring variant enzymes. *Journal of Pharmacology and Experimental Therapeutics*, 308(1), 300-6.
- Narayan, S., Tang, B., Head, S.R., Gilmartin, J., Sutcliffe, G., Dean, B., Thomas, E.A. 2008. Molecular profiles of schizophrenia in the CNS at different stages of illness. *Brain Research*, 1239, 235-48.
- Nasrallah, H.A. 2007. The case for long-acting antipsychotic agents in the post-CATIE era. *Acta Psychiatrica Scandinavica*, 115, 260-7.
- Nechanitzky, R., Akbas, D., Scherer, S., Györy, I., Hoyler, T., Ramamoorthy, S., *et al.* 2013. Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. *Nature Immunology*, 14, 867-75.
- Need, A., Keefe, R., Ge, D., Grossman, I., Dickson, S., McEvoy, J.P., Goldstein, D.B. 2009. Pharmacogenetics of antipsychotic response in the CATIE trial: a candidate gene analysis. *European Journal of Human Genetics*, 17(7), 946-57.
- Newcomer, J.W. 2005. Second-generation (atypical) antipsychotics and metabolic effects: a comprehensive literature review. *CNS Drugs*, 19 (S1), 1-93.
- Ni, X., Zhang, W., Huang, R.S. 2013. Pharmacogenomics discovery and implementation in genome-wide association studies era. *Wiley Interdisciplinary Reviews Systems Biology and Medicine*, 5, 1-9.
- Niemi, M., Schaeffeler, E., Lang, T., Fromm, M.F., Neuvonen, M., Kyrklund, C., *et al.* 2004. High plasma pravastatin concentrations are associated with single nucleotide polymorphisms and haplotypes of organic anion transporting polypeptide-C (OATP-C, *SLCO1B1*). *Pharmacogenetics*, 14(7), 429-40.
- Nyakutira, C., Röshammar, D., Chigutsa, E., Chonzi, P., Ashton, M., Nhachi, C., Masimirembwa, C. 2008. High prevalence of the *CYP2B6* 516G>T(*6) variant and effect on the population pharmacokinetics of efavirenz in HIV/AIDS outpatients in Zimbabwe. *European Journal of Clinical Pharmacology*, 64(4), 357-65.
- Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., Nakatani, Y. 1996. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell*, 87, 953-9.
- Oosthuizen, P., Emsley, R., Jadri, T.H., Keyter, N. 2004. A randomized, controlled comparison of the efficacy and tolerability of low and high doses of haloperidol in the treatment of first-episode psychosis. *International Journal of Neuropsychopharmacology*, 7(2), 125-31.
- Opler, M., Charap, J., Greig, A., Stein, V., Polito, S., Malaspina, D. 2013. Environmental risk factors and schizophrenia. *International Journal of Mental Health*, 42(1), 23-32.

- Overall, J.E., Gorham, D.R. 1962. The brief psychiatric rating scale. *Psychological Reports*, 10, 799–812.
- Ozomaro, U., Wahlstedt, C., Nemeroff, C.B. 2013. Personalized medicine in psychiatry: problems and promises. *BMC Medicine*, 11(132), 1-35.
- Pastinen, T. 2010. Genome-wide allele-specific analysis: insights into regulatory variation. *Nature Reviews Genetics*, 11(8), 533-8.
- Patel, V., Sumathipala, A. 2001. International representation in psychiatric literature: survey of six leading journals. *British Journal of Psychiatry*, 178(5), 406-9.
- Patel, V., Kleinman, A. 2003. Poverty and common mental disorders in developing countries. *Bulletin of the World Health Organization*, 81(8), 609-15.
- Patel, V., Belkin, G.S., Chockalingam, A., Cooper, J., Saxena, S., Unützer, J. 2013. Grand challenges: integrating mental health services into priority health care platforms. *PLoS Medicine*, 10(5). Available: <http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001448>.
- Peralta, V., Cuesta, M.J. 2003. The nosology of psychotic disorders: a comparison among competing classification systems. *Schizophrenia Bulletin*, 29, 413–25.
- Perkins, D., Lieberman, J., Gu, H., Tohen, M., McEvoy, J., Green, A., *et al.* 2004. Predictors of antipsychotic treatment response in patients with first-episode schizophrenia, schizoaffective and schizophreniform disorders. *British Journal of Psychiatry*, 185(1), 18-24.
- Petrovski, S., Fellay, J., Shianna, K.V., Carpenetti, N., Kumwenda, J., Kamanga, G., *et al.* 2011. Common human genetic variants and HIV-1 susceptibility: a genome-wide survey in a homogeneous African population. *AIDS*, 25(4), 513–18.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar D., R Core Team. 2014. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-117. Available: <http://CRAN.R-project.org/package=nlme>.
- Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., *et al.* 2014. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, 506(7487), 185-90.
- Qiu, W., Rogers, A.J., Damask, A., Raby, B.A., Klanderma, B.J., Duan, Q.L., *et al.* 2014. Pharmacogenomics: novel loci identification via integrating gene differential analysis and eQTL analysis. *Human Molecular Genetics*. Available: <http://hmg.oxfordjournals.org/content/early/2014/05/07/hmg.ddu191.short>.
- R Development Core Team. 2010. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available: <http://www.R-project.org>.

- Ramsay, M. 2012. Africa: continent of genome contrasts with implications for biomedical research and health. *FEBS Letters*, 586, 2813-9.
- Reith, W., Steimle, V., Mach, B. 1995. Molecular defects in the bare lymphocyte syndrome and regulation of MHC class II genes. *Immunology Today*, 16(11), 539-46.
- Remington, G., Foussias, G., Agid, O. 2010. Progress in defining optimal treatment outcome in schizophrenia. *CNS Drugs*, 24(1), 9-20.
- Reynolds, G.P. 2007. The impact of pharmacogenetics on the development and use of antipsychotic drugs. *Drug Discovery Today*, 12(21-22), 953-9.
- Rieder, M. 2014. Pharmacogenomics in children. In Yan, Q. (ed.). *Pharmacogenomics in Drug Discovery and Development*. 2nd Ed. New York: Humana Press. 687-708.
- Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kähler, A.K., Akterin, S., et al. 2013. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics*, 45(10), 1150-9.
- Ripke, S., Neale, B.M., Corvin, A., Walters, J.T.R., Farh, K-H., Holmans, P.A., et al. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511, 421-7.
- Ritchie, G.R.S., Dunham, I., Zeggini, E., Flicek, P. 2014. Functional annotation of noncoding sequence variants. *Nature Methods*, 11, 294-6.
- Robinson, D.G., Woerner, M.G., McMeniman, M., Mendelowitz, A., Bilder, R.M. 2004. Symptomatic and functional recovery from a first episode of schizophrenia or schizoaffective disorder. *American Journal of Psychiatry*, 161, 473-9.
- Rodriguez-Murillo, L., Gogos, J., Karayiorgou, M. 2012. The genetic architecture of schizophrenia: new mutations and emerging paradigms. *Annual Review of Medicine*, 63, 63-80.
- Samuels, D.C., Burn, D.J., Chinnery, P.F. 2009. Detecting new neurodegenerative disease genes: does phenotype accuracy limit the horizon? *Trends in Genetics*, 25(11), 486-8.
- Saxena, S., Skeen, S. 2012. No health without mental health: challenges and opportunities in global mental health. *African Journal of Psychiatry*, 15(6), 397-400.
- Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S., Snyder, M. 2012. Linking disease associations with regulatory information in the human genome. *Genome Research*, 22, 1748-59.
- Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., et al. 2010. Five-vertebrate ChIP-Seq reveals the evolutionary dynamics of transcription factor binding. *Science*, 328(5981), 1036-40.
- Schmitt, A., Malchow, B., Hasan, A., Falkai, P. 2014. The impact of environmental factors in severe psychiatric disorders. *Frontiers in Neuroscience*, 8(19), 1-32.
- Sham, P.C., Purcell, S.M. 2014. Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*, 15(5), 335-46.

- Shishodia, S., Aggarwal, B.B. 2004. Nuclear factor- κ B activation mediates cellular transformation, proliferation, invasion angiogenesis and metastasis of cancer. *Cancer Treatment and Research*, 119, 139-73.
- Sim, S.C., Risinger, C., Dahl, M.L., Aklillu, E., Christensen, M., Bertilsson, L., Ingelman-Sundberg, M. 2006. A common novel *CYP2C19* gene variant causes ultrarapid drug metabolism relevant for the drug response to proton pump inhibitors and antidepressants. *Clinical Pharmacology & Therapeutics*, 79(1), 103–13.
- Singh, S., Kumar, A., Agarwal, S., Phadke, S.R., Jaiswal, Y. 2014. Genetic insight of schizophrenia: past and future perspectives. *Gene*, 535(2), 97-100.
- Smolin, B., Karry, R., Gal-Ben-Ari, S., Ben-Shachar, D. 2012. Differential expression of genes encoding neuronal ion-channel subunits in major depression, bipolar disorder and schizophrenia: implications for pathophysiology. *International Journal of Neuropsychopharmacology*, 15, 869-82.
- Solé, X., Guinó, E., Valls, J., Iniesta, R., Moreno, V. 2006. SNPStats: a web tool for the analysis of association studies. *Bioinformatics*, 22(15), 1928-9.
- Stefansson, H., Ophoff, R.A., Steinberg, S., Ole, A., Andreassen, A., Cichon, S., *et al.* 2009. Common variants conferring risk of schizophrenia. *Nature*, 460(7256), 744-7.
- Stenson, P.D., Mort, M., Ball, E.V., Howells, K., Phillips, A.D., Thomas, N.S.T., Cooper, D.N. 2009. The Human Gene Mutation Database: 2008 update. *Genome Medicine*, 1(13). Available: <http://genomemedicine.com/content/1/1/13>.
- Strange, A., Riley, B.P., Spencer, C.C.A., Morris, D.W., Pirinen, M., O'Dushlaine, C.T., *et al.* 2012. Genome-wide association study implicates *HLA-C*01:02* as a risk factor at the major histocompatibility complex locus in schizophrenia. *Biological Psychiatry*, 72(8), 620-8.
- Strassnig, M., Miewald, J., Keshavan, M., Ganguli, R. 2007. Weight gain in newly diagnosed first-episode psychosis patients and healthy comparisons: one-year analysis. *Schizophrenia Research*, 93(1-3), 90–8.
- Stroup, T.S., McEvoy, J.P., Swartz, M.S., Byerly, M.J., Glick, I.D., Canive, J.M., *et al.* 2003. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development. *Schizophrenia Bulletin*, 29(1), 15-31.
- Sugatani, J., Yamakawa, K., Yoshinari, K., Machida, T., Takagi, H., Mori, M., *et al.* 2002. Identification of a defect in the *UGT1A1* gene promoter and its association with hyperbilirubinemia. *Biochemical and Biophysical Research Communications*, 292(2), 492–7.
- Sullivan, P.F., Kendler, K.S., Neale, M.C. 2003. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Archives of General Psychiatry*, 60, 1187-92.

- Sullivan, P.F. 2005. The genetics of schizophrenia. *PLoS Medicine*, 2(7). Available: <http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.0020212>.
- Sullivan, P.F. 2010. The Psychiatric GWAS Consortium: big science comes to psychiatry. *Neuron*, 68(2), 182-6.
- Sullivan, P.F., Daly, M., O'Donovan, M. 2012. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature Reviews Genetics*, 13(8), 537-51.
- Sun, J., Xu, H., Zhao, Z. 2012. Network-assisted investigation of antipsychotic drugs and their targets. *Chemistry & Biodiversity*, 9, 900-10.
- Suzuki, T., Remington, G., Mulsant, B.H., Uchida, H., Rajji, T.K., Graff-Guerrero, A., *et al.* 2012. Defining treatment-resistant schizophrenia and response to antipsychotics: a review and recommendation. *Psychiatry Research*, 197(1-2), 1-6.
- Tandon, R., Keshavan, M., Nasrallah, H. 2008. Schizophrenia, "just the facts" what we know in 2008. 2. Epidemiology and etiology. *Schizophrenia Research*, 102(1-3), 1-18.
- Tandon, R., Keshavan, M., Nasrallah, H. 2010. Schizophrenia, "just the facts" 5. Treatment and prevention. Past, present, and future. *Schizophrenia Research*, 122(1-3), 1-23.
- Tandon, R., Nasrallah, H., Keshavan, M. 2009. Schizophrenia, "just the facts" 4. Clinical features and conceptualization. *Schizophrenia Research*, 110(1-3), 1-23.
- Tandon, R. 2012. The nosology of schizophrenia: toward DSM-5 and ICD-11. *Psychiatric Clinics of North America*, 35(3), 557-69.
- Tandon, R. 2014. Schizophrenia and other psychotic disorders in diagnostic and statistical manual of mental disorders (DSM)-5: clinical implications of revisions from DSM-IV. *Indian Journal of Psychological Medicine*, 36(3), 223-5.
- Tarricone, I., Ferrari Gozzi, B., Serretti, A., Grieco, D., Berardi, D. 2010. Weight gain in antipsychotic-naïve patients: a review and meta-analysis. *Psychological Medicine*, 40(2), 187-200.
- Thorn, C.F., Klein, T.E., Altman, R.B. 2010. Pharmacogenomics and bioinformatics: PharmGKB. *Pharmacogenomics*, 11(4), 501-5.
- Tomlinson, M., Rudan, I., Saxena, S., Swartz, L., Tsai, A.C., Patel, V. 2009. Setting priorities for global mental health research. *Bulletin of the World Health Organization*, 87(6), 438-46.
- Tomppo, L., Hennah, W., Lahermo, P., Loukola, A., Tuulio-Henriksson, A., Suvisaari, J., *et al.* 2009 Association between genes of disrupted in schizophrenia 1 (*DISC1*) interactors and schizophrenia supports the role of the *DISC1* pathway in the etiology of major mental illness. *Biological Psychiatry*, 65(12), 1055-62.

- Tourjman, V., Kouassi, É., Koué, M-É., Rocchetti, M., Fortin-Fournier, S., Fusar-Poli, P., Potvin, S. 2013. Antipsychotics' effects on blood levels of cytokines in schizophrenia: a meta-analysis. *Schizophrenia Research*, 151(1-3), 43-7.
- Tsuang, M.T., Bar, J.L., Stone, W.S., Faraone, S.V. 2004. Gene-environment interactions in mental disorders. *World Psychiatry*, 3, 73–83.
- van Dongen, J., Boomsma, D.I. 2013. The evolutionary paradox and the missing heritability of schizophrenia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 162B(2), 122-36.
- van Os, J., Kapur, S. 2009. Schizophrenia. *The Lancet*, 374(9690), 635-45.
- van Os, J., Kenis, G., Rutten, B.P. 2010. The environment and schizophrenia. *Nature*, 468(7321), 203-12.
- Vogel, F. 1959. Moderne Probleme der Humangenetik. In Heilmeyer, L., Schoen, R., de Rudder, B. (Eds.). *Ergebnisse der Inneren Medizin und Kinderheilkunde*. Springer Berlin Heidelberg, 52-125.
- Wang, P.S., Lane, M., Olfson, M., Pincus, H.A., Wells, K.B., Kessler, R.C. 2005. Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62, 629–40.
- Wang, D., Guo, Y., Wrighton, S.A., Cooke, G.E., Sadee, W. 2010. Intronic polymorphism in *CYP3A4* affects hepatic expression and response to statin drugs. *Pharmacogenomics*, 11(4), 274-86.
- Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., *et al.* 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*, 38. Available: http://nar.oxfordjournals.org/content/38/suppl_2/W214.full.
- Warnich, L., Drögemöller, B.I., Pepper, M.S., Dandara, C., Wright, G.E.B. 2011. Pharmacogenomic research in South Africa: lessons learned and future opportunities in the Rainbow Nation. *Current Pharmacogenomics and Personalized Medicine*, 9(3), 191-207.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., *et al.* 2013. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*, 42. Available: <http://nar.oxfordjournals.org/content/42/D1/D1001.full>.
- Whiteford, H.A., Degenhardt, L., Rehm, J., Baxter, A.J., Ferrari, A.J., Erskine, H.E. 2013. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet*, 382, 1575-86.
- Wirgenes, K.V., Sønderby, I.E., Haukvik, U.K., Mattingsdal, M., Tesli, M., Athanasiu, L., *et al.* 2012. *TCF4* sequence variants and mRNA levels are associated with

- neurodevelopmental characteristics in psychotic disorders. *Translational Psychiatry*, 2. Available: <http://www.nature.com/tp/journal/v2/n5/full/tp201239a.html>.
- Woodward, T.S., Jung, K., Smith, G.N., Hwang, H., Barr, A.M., Procyshyn, R.M., *et al.* 2014. Symptom changes in five dimensions of the Positive and Negative Syndrome Scale in refractory psychosis. *European Archives of Psychiatry and Clinical Neuroscience*, 264(8), 673-82.
- World Health Organization. 2008. *The global burden of disease: 2004 update*. Geneva: World Health Organization.
- World Health Organization. 2011a. *Mental Health Atlas 2011*. Geneva: World Health Organization.
- World Health Organization. 2011b. *Global status report on noncommunicable diseases 2010*. Geneva: World Health Organization.
- World Health Organization. 2015. *The International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10) Version:2015*. Geneva: World Health Organization.
- Wright, J. 2014. Genetics: unravelling complexity. *Nature*, 508, S6-7.
- Xu, B., Roos, J.L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., *et al.* 2011. Exome sequencing supports a *de novo* mutational paradigm for schizophrenia. *Nature Genetics*, 43, 864–8.
- Yu, W., Gwinn, M., Clyne, M., Yesupriya, A., Khoury, M.J. 2008. A navigator for human genome epidemiology. *Nature Genetics*, 40(2), 124-5.
- Yuan, H.Y., Chen, J.J., Lee, M.T., Wung, J.C., Chen, Y.F., Charng, M.J., *et al.* 2005. A novel functional *VKORC1* promoter polymorphism is associated with inter-individual and inter-ethnic differences in warfarin sensitivity. *Human Molecular Genetics*, 14(13), 1745–51.
- Zandi, P.P., Judy, J.T. 2010. The promise and reality of pharmacogenetics in psychiatry. *Psychiatric Clinics of North America*, 33, 181-224.
- Zanger, U.M., Fischer, J., Raimundo, S., Stüven, T., Evert, B.O., Schwab, M., Eichelbaum, M. 2001. Comprehensive analysis of the genetic factors determining expression and function of hepatic CYP2D6. *Pharmacogenetics*, 11(7), 573–85.
- Zanger, U.M., Raimundo, S., Eichelbaum, M. 2004. Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 369, 23-37.
- Zhang, J-P., Malhotra, A.K. 2013a. Genetics of schizophrenia: what do we know? *Current Psychiatry*, 12(3), 24-33.

- Zhang, J-P., Malhotra, A.K. 2013b. Pharmacogenetics of antipsychotics: recent progress and methodological issues. *Expert Opinion on Drug Metabolism & Toxicology*, 9(2), 183-91.
- Zhao, H., Nettleton, D., Soller, M., Dekkers, J.C.M. 2005. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genetical Research*, 86, 77-87.

ELECTRONIC SOURCES

DAVID version 6.7

<http://david.abcc.ncifcrf.gov/home.jsp>

Accessed September 2014

ENCODE: UCSC Genome Browser

<http://genome.ucsc.edu/ENCODE/>

Accessed October 2013

Ensembl Genome Browser release 73

<http://www.ensembl.org/index.html>

Accessed October 2013

FANTOM5

<http://fantom.gsc.riken.jp/5/sstar>

Accessed September 2014

FDA Table of Pharmacogenomic Biomarkers in Drug Labeling

<http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm>

Accessed January 2015

GeneMANIA

<http://www.genemania.org/>

Accessed September 2014

HGNC

<http://www.genenames.org/>

Accessed February 2014

HuGE Navigator GWAS Integrator version 2.0

<http://hugenavigator.net/HuGENavigator/home.do>

Accessed January 2014

Jaspar version 5.0 ALPHA

<http://jaspar.genereg.net/>

Accessed January 2014

NHGRI GWAS Catalog

<http://www.genome.gov/gwastudies/>

Accessed January 2014

Interactive NHGRI GWAS Catalog

<http://www.ebi.ac.uk/fgpt/gwas/>

Accessed January 2014

OPTiMiSE

<http://www.optimisetrials.eu/>

Accessed April 2015

PharmGKB

<http://www.pharmgkb.org/>

Accessed July 2013

PolyPhen-2

<http://genetics.bwh.harvard.edu/pph2/>

Accessed July 2013

PubMed

<http://www.ncbi.nlm.nih.gov/pubmed>

Accessed May 2013

RegulomeDB

<http://regulome.stanford.edu/>

Accessed October 2013

rSNPBase

<http://rsnp.psych.ac.cn/>

Accessed January 2014

SeattleSeq Annotation 137 version 8.07

<http://snp.gs.washington.edu/SeattleSeqAnnotation137/>

Accessed February 2014

SNAP version 2.2

<http://www.broadinstitute.org/mpg/snap/>

Accessed June 2013

SNP Genotyping Assay Search Tool

<http://www.lifetechnologies.com/za/en/home/life-science/pcr/real-time-pcr/real-time-pcr-assays/snp-genotyping-taqman-assays.html>

Accessed April 2014

SNPStats

<http://bioinfo.iconcologia.net/snpstats/start.htm>

Accessed November 2014

sTRAP

http://trap.molgen.mpg.de/cgi-bin/trap_two_seq_form.cgi

Accessed December 2013

tRap download

<http://trap.molgen.mpg.de/cgi-bin/download.cgi>

Accessed December 2013

APPENDIX A: Bioinformatics supplementary data

Script S1: Unix shell commands for TRAP workflow, including tRap (<http://trap.molgen.mpg.de/cgi-bin/download.cgi>), written by Dr N. Ishaque of the German Cancer Research Center (Heidelberg, Germany)*.

```
tr '\n' '|' < AllSNPs.txt > AllSNPs2.txt

egrep -w 'rsID1|rsID2|rsID3|rsIDn-1|rsIDn' 00-All.SNV.bed > All_results.txt

awk '{print $1"\t"$2"\t"($3+1)"\t"$4"\t"$5"\t"$6}' All_results.txt > All_results.bed

bedtools getfasta -fi GRCh37_FINAL.fa -bed All_results.bed -fo Allfasta

awk '{print $1"\t"($2-16)"\t"($3-1)"\t"$4}' All_results.bed > All_pre15.bed

awk '{print $1"\t"($2)"\t"($3+15)"\t"$4}' All_results.bed > All_post15.bed

bedtools getfasta -fi GRCh37_FINAL.fa -bed All_pre15.bed -fo All_pre15.fa

bedtools getfasta -fi GRCh37_FINAL.fa -bed All_post15.bed -fo All_post15.fa

perl weave_fasta_for_sTRAP.pl All_results.txt All_pre15.fa All_post15.fa >
All_results_weaved_for_sTRAP.fa

grep WT All_results_weaved_for_sTRAP.fa > AllpairsWT.txt

grep MUT All_results_weaved_for_sTRAP.fa > AllpairsMUT.txt

paste AllpairsWT.txt AllpairsMUT.txt > Allpairs.txt

sed 's/>/|g' Allpairs.txt > Allpairs.pairs

R

library(tRap)
data(jaspar)
library(Biostrings)
sequences = readFASTA("All_results_weaved_for_sTRAP.fa")
names(sequences) = gsub(">", "", sapply(sequences, "[[", "desc"))
pairs = read.csv("Allpairs.pairs", sep="\t", header=F, stringsAsFactors=F)
strap = rank.factors.for.pairs(jaspar, sequences, pairs)
filtered = strap[which(strap[, "min.p"] < 0.01 & abs(strap[, "log.ratio"]) > log(1.2)),]
save.image(file="strap_try1.RData")
write.table(filtered, file="All_strap_results_for_pairs.txt", sep="\t", quote=F, row.names=F)

q()

cut -f 3 All_strap_results_for_pairs.txt | sed 's|,|\n|g' | sort | uniq -c | sort -rg >
All_strap_results_for_pairs_u.txt

- Manually search for motifs on Jaspar database ("All_strap_results_for_pairs_u_hs.txt")
- Create .txt file with list of Jaspar IDs for motifs only in humans

grep -f All_strap_results_for_pairs_u_hs.txt All_strap_results_for_pairs.txt >
All_strap_results_for_pairs_onlyhs.txt
```

* Input SNPs were obtained from HuGE Navigator, the NHGRI Catalog, and subsequent LD analysis, and saved as "AllSNPs.txt" prior to analysis. The "control" traits were each analysed in the same way, with input files "Control1SNPs.txt" for response to antidepressants, "Control2SNPs.txt" for response to hepatitis C treatment, and "Control3SNPs.txt" for eye colour. Italicised text represents commands of tRap script. Editable fields are indicated in bold. Instructions are indicated in grey.

Script S2: Perl script "weave_fasta_for_sTRAP.pl" written by Dr N. Ishaque (German Cancer Research Center, Heidelberg, Germany).

```
# use strict;
```

```
my $usage = "This program weaves fasta files for sTRAP\n\n\t$0 [dbSNP bed file, 4th col = ref, 5th col = alt] [PRE FASTA] [POST FASTA]\n\n";
my $nex = shift or die "Please provide SNP CSV file\n\n$usage";
my $pre = shift or die "Please provide PRE FASTA file\n\n$usage";
my $post = shift or die "Please provide POST FASTA file\n\n$usage";
open (NEX_F, "$nex") or die "Cannot open NEXUS FILE '$nex'\n\n$usage";
open (PRE_F, "$pre") or die "Cannot open PRE FASTA FILE '$nex'\n\n$usage";
open (POST_F, "$post") or die "Cannot open POST FASTA FILE '$nex'\n\n$usage";
```

```
my %iupac=("B" => ["C","G","T"], "D" => ["A","G","T"], "K" => ["G","T"], "M" => ["A","C"], "R" => ["A","G"], "S" => ["C","G"], "V" => ["A","C","G"], "W" => ["A","T"], "Y" => ["C","T"]);
```

```
while (<NEX_F>){ if (/^(.*)\t(.*)\t(.*)\t(.*)\t(.*)$/){my ($snp, $ref, $alt) = ($4,$5,$6);
my $pre_f_line= <PRE_F>;
$pre_f_line= <PRE_F>;
my $post_f_line= <POST_F>;
$post_f_line= <POST_F>;
chomp ($pre_f_line);
chomp ($post_f_line);
if ($alt =~ m/A/ || $alt =~ m/C/ || $alt =~ m/G/ || $alt =~ m/T/ )
{
print ">$snp"."_WT\n$pre_f_line$ref$post_f_line\n";
print ">$snp"."_MUT\n$pre_f_line$alt$post_f_line\n";
}
else {
#my @bases=$iupac{$alt};
foreach my $base (@{$iupac{$alt}}){
# warn "$snp $alt $base\n";
print ">$snp"."_$alt$base"."_WT\n$pre_f_line$ref$post_f_line\n";
print ">$snp"."_$alt$base"."_MUT\n$pre_f_line$base$post_f_line\n";
}
}
}
else {
die "invalid line in SNP NEXUS file: $_\n";
}
}
close (NEX_F); close (PRE_F); close (POST_F);
```


Box S1: FANTOM5 cell lines for brain and liver tissue, used to assess gene expression levels (<http://fantom.gsc.riken.jp/5/sstar>).

Brain tissue

- amygdala - adult, donor10196.CNhs13793.10167-103B5
- amygdala, adult, donor10252.CNhs12311.10151-102I7
- Astrocyte - cerebellum, donor1.CNhs11321.11500-119F6
- Astrocyte - cerebellum, donor3.CNhs12117.11661-122F5
- Astrocyte - cerebellum, donor2.CNhs12081.11580-120F5
- Astrocyte - cerebral cortex, donor1.CNhs10864.11235-116D2
- Astrocyte - cerebral cortex, donor2.CNhs11960.11316-117D2
- Astrocyte - cerebral cortex, donor3.CNhs12005.11392-118C6
- brain, fetal, pool1.CNhs11797.10085-102B4
- brain, adult, pool1.CNhs10617.10012-101C3
- brain, adult, donor1.CNhs11796.10084-102B3
- cerebellum - adult, donor10196.CNhs13799.10173-103C2
- cerebellum, adult, donor10252.CNhs12323.10166-103B4
- cerebellum, adult, pool1.CNhs11795.10083-102B2
- cerebral meninges, adult.CNhs12840.10188-103D8
- corpus callosum, adult, pool1.CNhs10649.10042-101F6
- diencephalon, adult.CNhs12610.10193-103E4
- dura mater, adult, donor1.CNhs10648.10041-101F5
- frontal lobe, adult, pool1.CNhs10647.10040-101F4
- globus pallidus - adult, donor10196.CNhs13801.10175-103C4
- globus pallidus, adult, donor10252.CNhs12319.10161-103A8
- hippocampus - adult, donor10196.CNhs13795.10169-103B7
- hippocampus, adult, donor10252.CNhs12312.10153-102I9
- locus coeruleus, adult, donor10252.CNhs12322.10165-103B3
- locus coeruleus - adult, donor10196.CNhs13808.10182-103D2
- medial frontal gyrus - adult, donor10196.CNhs13796.10170-103B8
- medial temporal gyrus - adult, donor10196.CNhs13809.10183-103D3
- medial temporal gyrus, adult, donor10252.CNhs12310.10150-102I6
- medulla oblongata, adult, donor10252.CNhs12315.10155-103A2
- medulla oblongata, adult, pool1.CNhs10645.10038-101F2
- medulla oblongata - adult, donor10196.CNhs13800.10174-103C3
- Meningeal Cells, donor1.CNhs11320.11493-119E8
- Meningeal Cells, donor2.CNhs12080.11573-120E7
- Meningeal Cells, donor3.CNhs12731.11654-122E7
- middle temporal gyrus, donor10252.CNhs12316.10156-103A3
- nucleus accumbens, adult, pool1.CNhs10644.10037-101F1
- occipital cortex, adult, donor10252.CNhs12320.10163-103B1
- occipital cortex - adult, donor10196.CNhs13798.10172-103C1
- occipital lobe, adult, donor1.CNhs11787.10076-102A4
- occipital lobe, fetal, donor1.CNhs11784.10073-102A1
- paracentral gyrus, adult, pool1.CNhs10642.10035-101E8
- parietal lobe, adult, pool1.CNhs10641.10034-101E7
- parietal lobe, adult, donor10252.CNhs12317.10157-103A4
- parietal lobe, fetal, donor1.CNhs11782.10072-101I9
- parietal lobe - adult, donor10196.CNhs13797.10171-103B9
- pineal gland - adult, donor10196.CNhs13804.10179-103C8
- pineal gland, adult, donor10252.CNhs12228.10160-103A7
- pituitary gland - adult, donor10196.CNhs13805.10180-103C9
- pons, adult, pool1.CNhs10640.10033-101E6
- postcentral gyrus, adult, pool1.CNhs10638.10032-101E5
- putamen, adult, donor10196.CNhs12324.10176-103C5
- Smooth Muscle Cells - Brain Vascular, donor1.CNhs10863.11234-116D1
- Smooth Muscle Cells - Brain Vascular, donor2.CNhs11900.11315-117D1

Brain tissue (continued)

- Smooth Muscle Cells - Brain Vascular, donor3.CNhs12004.11391-118C5
- substantia nigra, adult, donor10252.CNhs12318.10158-103A5
- temporal lobe, adult, pool1.CNhs10637.10031-101E4
- temporal lobe, fetal, donor1, tech_rep2.CNhs12996.10063-101H9
- temporal lobe, fetal, donor1, tech_rep1.CNhs11772.10063-101H9
- thalamus, adult, donor10252.CNhs12314.10154-103A1
- thalamus - adult, donor10196.CNhs13794.10168-103B6

Liver tissue

- Hepatic Sinusoidal Endothelial Cells, donor1.CNhs12075.11521-119H9
- Hepatic Sinusoidal Endothelial Cells, donor2.CNhs12092.11601-120H8
- Hepatic Stellate Cells (lipocyte), donor1.CNhs11335.11524-119I3
- Hepatic Stellate Cells (lipocyte), donor2.CNhs12093.11604-120I2
- Hepatocyte, donor1.CNhs12340.11523-119I2
- Hepatocyte, donor2.CNhs12349.11603-120I1
- Hepatocyte, donor3.CNhs12626.11684-122I1
- liver, fetal, pool1.CNhs11798.10086-102B5
- liver, adult, pool1.CNhs10624.10018-101C9
- Mesenchymal stem cells - hepatic, donor0.CNhs10845.11218-116B3
- Mesenchymal Stem Cells - hepatic, donor2.CNhs12730.11618-122A7

Table S1: Unique tRap motifs and number of associated SNPs for antipsychotic response and three control traits.

Antipsychotic response			Control 1: Antidepressant response		Control 2: Hepatitis C response		Control 3: Eye colour	
Jaspar ID	SNPs	Motif name	Jaspar ID	SNPs	Jaspar ID	SNPs	Jaspar ID	SNPs
MA0156.1	2	FEV	MA0197.1	6	MA0393.1	1	MA0268.1	2
MA0107.1	2	RELA	MA0151.1	6	MA0389.1	1	MA0446.1	1
MA0101.1	2	REL	MA0398.1	5	MA0321.1	1	MA0441.1	1
MA0080.2	2	SPI1	MA0387.1	5	MA0302.1	1	MA0434.1	1
MA0059.1	2	MYC::MAX	MA0346.1	5			MA0431.1	1
MA0259.1	1	HIF1A::ARNT	MA0231.1	5			MA0429.1	1
MA0160.1	1	NR4A2	MA0038.1	5			MA0428.1	1
MA0112.2	1	ESR1	MA0243.1	4			MA0425.1	1
MA0091.1	1	TAL1::TCF3	MA0200.1	4			MA0424.1	1
MA0076.1	1	ELK4	MA0125.1	4			MA0381.1	1
MA0037.1	1	GATA3	MA0102.2	4			MA0380.1	1
MA0031.1	1	FOXD1	MA0032.1	4			MA0367.1	1
			MA0457.1	3			MA0361.1	1
			MA0455.1	3			MA0358.1	1
			MA0413.1	3			MA0353.1	1
			MA0407.1	3			MA0339.1	1
			MA0284.1	3			MA0338.1	1
			MA0257.1	3			MA0337.1	1
			MA0254.1	3			MA0326.1	1
			MA0251.1	3			MA0292.1	1
			MA0248.1	3			MA0285.1	1
			MA0241.1	3			MA0280.1	1
			MA0237.1	3			MA0277.1	1
			MA0236.1	3			MA0270.1	1
			MA0235.1	3			MA0262.1	1
			MA0230.1	3			MA0233.1	1
			MA0229.1	3			MA0090.1	1
			MA0225.1	3			MA0078.1	1
			MA0224.1	3			MA0041.1	1
			MA0223.1	3				
			MA0220.1	3				
			MA0206.1	3				
			MA0202.1	3				
			MA0198.1	3				
			MA0195.1	3				
			MA0187.1	3				
			MA0184.1	3				
			MA0183.1	3				
			MA0181.1	3				
			MA0179.1	3				
			MA0178.1	3				
			MA0177.1	3				
			MA0175.1	3				
			MA0172.1	3				
			MA0167.1	3				
			MA0132.1	3				

83 more

Table S2: 118 affected genes uploaded to DAVID and GeneMANIA for further analyses.

Gene symbol	Gene name
<i>ACTG1</i>	actin, gamma 1
<i>ADH7</i>	alcohol dehydrogenase 7 (class IV), mu/sigma polypeptide
<i>AK7</i>	adenylate kinase 7
<i>ALPK2</i>	alpha-kinase 2
<i>ANKS1B</i>	ankyrin repeat and sterile alpha motif domain containing 1B
<i>APOA1BP</i>	apolipoprotein A-I binding protein
<i>ARFGEF2</i>	ADP-ribosylation factor guanine nucleotide-exchange factor 2
<i>ARRDC3</i>	arrestin domain containing 3
<i>ASTN2</i>	astrotactin 2
<i>ATP1A2</i>	ATPase, Na ⁺ /K ⁺ transporting, alpha 2 (+) polypeptide
<i>AVL9</i>	AVL9 homolog (<i>S. cerevisiae</i>)
<i>BANK1</i>	B-cell scaffold protein with ankyrin repeats 1
<i>BCL7B</i>	B-cell CLL/lymphoma 7B
<i>C18orf21</i>	chromosome 18 open reading frame 21
<i>C19Orf12</i>	chromosome 19 open reading frame 12
<i>C1orf194</i>	chromosome 1 open reading frame 194
<i>C2orf34</i>	chromosome 2 open reading frame 34
<i>C6orf25</i>	chromosome 6 open reading frame 25
<i>C8orf56</i>	chromosome 8 open reading frame 56
<i>C9orf37</i>	chromosome 9 open reading frame 37
<i>CACNB1</i>	calcium channel, voltage-dependent, beta 1 subunit
<i>CASQ1</i>	calsequestrin 1 (fast-twitch, skeletal muscle)
<i>CCBE1</i>	collagen and calcium binding EGF domains 1
<i>CCDC114</i>	coiled-coil domain containing 114
<i>CCNG2</i>	cyclin G2
<i>CCT5</i>	chaperonin containing TCP1, subunit 5 (epsilon)
<i>CD74</i>	CD74 molecule, major histocompatibility complex, class II invariant chain
<i>CISD2</i>	CDGSH iron sulfur domain 2
<i>CKAP4</i>	cytoskeleton-associated protein 4
<i>CKS2</i>	CDC28 protein kinase regulatory subunit 2
<i>CLMN</i>	calmin (calponin-like, transmembrane)
<i>CNTNAP5</i>	contactin associated protein-like 5
<i>COL9A2</i>	collagen, type IX, alpha 2
<i>CRLS1</i>	cardiolipin synthase 1
<i>CUL5</i>	cullin 5
<i>DALRD3</i>	DALR anticodon binding domain containing 3
<i>DCAF8</i>	WD repeat domain 42A
<i>DDX11</i>	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 11
<i>DICER1</i>	dicer 1, ribonuclease type III
<i>DRD2</i>	dopamine receptor D2
<i>EFHB</i>	EF-hand domain family, member B
<i>EFNA4</i>	ephrin-A4
<i>EFR3B</i>	EFR3 homolog B

Gene symbol	Gene name
<i>EHF</i>	ets homologous factor
<i>EIF4E2</i>	eukaryotic translation initiation factor 4E family member 2
<i>EMP3</i>	epithelial membrane protein 3
<i>ENPP6</i>	ectonucleotide pyrophosphatase/phosphodiesterase 6
<i>FAM173B</i>	family with sequence similarity 173, member B
<i>FAM55B</i>	family with sequence similarity 55, member B
<i>FHOD3</i>	formin homology 2 domain containing 3
<i>GABRA5</i>	gamma-aminobutyric acid (GABA) A receptor, alpha 5
<i>GLS2</i>	glutaminase 2 (liver, mitochondrial)
<i>GPR137B</i>	G protein-coupled receptor 137B
<i>GPR98</i>	G protein-coupled receptor 98
<i>HBD</i>	hemoglobin, delta
<i>HEATR4</i>	HEAT repeat containing 4
<i>HSD11B1L</i>	hydroxysteroid (11-beta) dehydrogenase 1-like
<i>ING2</i>	inhibitor of growth family, member 2
<i>KIAA1549</i>	KIAA1549
<i>LAMP3</i>	lysosomal-associated membrane protein 3
<i>LAYN</i>	layilin
<i>LRRN4CL</i>	LRRN4 C-terminal like
<i>LSM5</i>	LSM5 homolog, U6 small nuclear RNA associated
<i>MANBA</i>	mannosidase, beta A, lysosomal
<i>MAPKSP1</i>	MAPK scaffold protein 1
<i>MEIS2</i>	Meis homeobox 2
<i>MIR582</i>	microRNA 582
<i>MYOG</i>	myogenin (myogenic factor 4)
<i>NDUFS2</i>	NADH dehydrogenase (ubiquinone) Fe-S protein 2
<i>NFKB1</i>	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1
<i>NHEDC1</i>	similar to Na ⁺ /H ⁺ exchanger domain containing 1
<i>NHEDC2</i>	Na ⁺ /H ⁺ exchanger domain containing 2
<i>OSBPL10</i>	oxysterol binding protein-like 10
<i>PANK3</i>	pantothenate kinase 3
<i>PCP4L1</i>	Purkinje cell protein 4 like 1
<i>PDE4D</i>	cAMP-specific phosphodiesterase 4D
<i>PEG10</i>	paternally expressed 10
<i>PIGK</i>	phosphatidylinositol glycan anchor biosynthesis, class K
<i>PKIG</i>	protein kinase (cAMP-dependent, catalytic) inhibitor gamma
<i>PKN2</i>	protein kinase N2
<i>PLEKHA4</i>	pleckstrin homology domain containing, family A, member 4
<i>PPA2</i>	pyrophosphatase (inorganic) 2
<i>PPARD</i>	peroxisome proliferator-activated receptor delta
<i>PPP1R9B</i>	protein phosphatase 1, regulatory (inhibitor) subunit 9B
<i>PSG6</i>	pregnancy specific beta-1-glycoprotein 6
<i>PTS</i>	6-pyruvoyltetrahydropterin synthase
<i>PXDN</i>	peroxidasin homolog
<i>RAB25</i>	RAB25, member RAS oncogene family

Gene symbol	Gene name
<i>RB1</i>	retinoblastoma 1
<i>RBM7</i>	RNA binding motif protein 7
<i>RNF144A</i>	ring finger protein 144A
<i>SFRS5</i>	splicing factor, arginine/serine-rich 5
<i>SFTPC</i>	surfactant protein C
<i>SH2D3C</i>	SH2 domain containing 3C
<i>SHC1</i>	SHC (Src homology 2 domain containing) transforming protein 1
<i>SLAMF1</i>	signaling lymphocytic activation molecule family member 1
<i>SLC22A23</i>	solute carrier family 22, member 23
<i>SLC26A9</i>	solute carrier family 26, member 9
<i>SLC39A8</i>	solute carrier family 39 (zinc transporter), member 8
<i>SMTNL2</i>	smoothelin-like 2
<i>SNORD116-27</i>	small nucleolar RNA, C/D box 116-27
<i>SOCS3</i>	suppressor of cytokine signaling 3
<i>SPOPL</i>	speckle-type POZ protein-like
<i>TC2N</i>	tandem C2 domains, nuclear
<i>THRA</i>	thyroid hormone receptor, alpha
<i>TJP1</i>	tight junction protein 1 (zona occludens 1)
<i>TMED2</i>	transmembrane emp24 domain trafficking protein 2
<i>TOMM40L</i>	translocase of outer mitochondrial membrane 40 homolog
<i>VILL</i>	villin-like
<i>WDR27</i>	WD repeat domain 27
<i>WDR74</i>	WD repeat domain 74
<i>ZBTB7B</i>	zinc finger and BTB domain containing 7B
<i>ZFAND3</i>	zinc finger, AN1-type domain 3
<i>ZFYVE26</i>	zinc finger, FYVE domain containing 26
<i>ZNF202</i>	zinc finger protein 202
<i>ZNF276</i>	zinc finger protein 276
<i>ZNF74</i>	zinc finger protein 74
<i>ZNF860</i>	zinc finger protein 860

APPENDIX B: Association analyses supplementary data**Table S3:** SNPs genotyped by TaqMan® OpenArray® in the FES cohort.

SNP	Alleles		Assay ID
	Major ^a	Minor ^a	
rs10458561	G	A	C__3184113_10
rs6688363	C	T	C__9406882_10
rs10218843	A	G	C__2823177_10
rs11265461	T	C	C__2823180_10
rs6427540	C	T	C_29332596_10
rs7520258	T	C	C_27150348_10
rs6741819	C	T	C_29310709_10
rs17727261	C	T	C_25927585_20
rs62161711	G	A	AHWSJXV
rs10170310	A	G	C__30363993_10
rs747559	G	A	C__804250_10
rs3774933	T	C	C_27479814_10
rs1599961	A	G	C__8935034_10
rs230534	C	T	C__3066477_10
rs230532	A	T	C__3066475_10
rs230529	T	C	C__804246_10
rs230526	A	G	C__804243_10
rs118882	C	T	C_176056349_10
rs230520	A	G	C__3066470_10
rs230505	G	T	C__3066462_10
rs230504	C	T	C__804227_10
rs230492	G	A	C__3066459_10
rs230493	T	A	C__3066458_10
rs230495	G	A	C__3066455_10
rs230539	A	G	C__804223_10
rs3774959	G	A	C_26458339_10
rs4648055	G	A	C__3066440_10
rs17440909	C	T	C_34264764_10
rs17742120	A	G	C_34264798_10
rs1352318	G	A	C__8339393_10
rs10492354	G	A	C_30433133_20

^a According to frequencies within the FES cohort.

APPENDIX B

ASSOCIATION ANALYSES SUPPLEMENTARY DATA

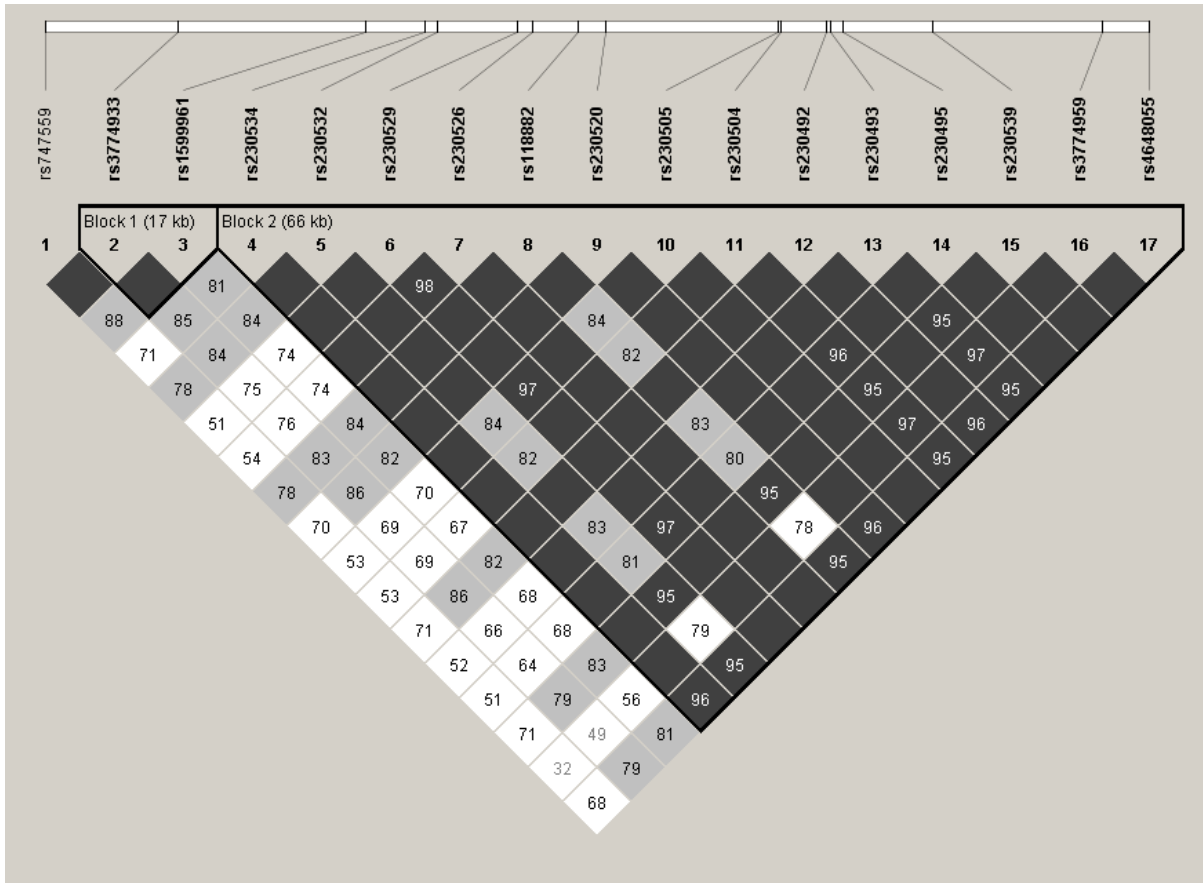


Figure S1: Two haplotype blocks on chromosome four, designated by Haploview version 4.2 ($D' > 0.7 - > 0.98$) (Barrett *et al.*, 2005). Dark squares indicate significant LD between SNPs; numbers within squares represent D' values as percentages.

Tables S4a – e: Haplotypes in the FES cohort from Haploview with frequencies ≥ 0.01 .

S4a: Chromosome 1

rs10218843	rs11265461	rs6427540	Frequency
A	T	C	0.61
G	C	T	0.26
G	C	C	0.13

S4b: Chromosome 2

rs62161711	rs10170310	Frequency
G	A	0.72
A	G	0.27

S4c: Chromosome 4 block 1

rs3774933	rs1599961	Frequency
C	A	0.48
T	G	0.47
T	A	0.05

S4d: Chromosome 4 block 2

rs230534	rs230532	rs230529	rs230526	rs118882	rs230520	rs230505	rs230504	rs230492	rs230493	rs230495	rs230539	rs3774959	rs4648055	Frequency
C	A	C	G	C	A	T	C	G	T	G	A	G	G	0.46
T	T	T	A	T	G	G	T	A	A	A	G	A	A	0.20
C	T	T	A	T	A	G	T	G	A	A	A	A	G	0.10
C	A	T	A	C	A	G	C	G	T	A	A	G	G	0.09
C	A	T	A	C	A	G	C	G	A	A	A	A	G	0.04
C	A	T	A	C	A	G	T	G	T	A	A	A	G	0.03
C	T	T	A	T	A	T	C	G	T	G	A	G	G	0.02
C	T	T	A	T	A	G	C	G	T	A	A	G	G	0.02

S4e: Chromosome 5

rs17440909	rs17742120	Frequency
C	A	0.85
T	G	0.15

Table S5: Significant associations ($P < 0.05$) with treatment outcomes in the FES cohort. Highlighted associations survived Bonferroni correction.

Response measurement	Variant/ haplotype	Association	P-value
PANSS Positive	rs230504	Allelic	0.0344
PANSS Positive	rs230493	Allelic	0.0249
PANSS Positive	rs230495	Allelic	0.0399
PANSS Positive	rs3774959	Genotypic	0.0207
PANSS Positive		Allelic	0.0112
PANSS Positive	C.A.T.A.C.A.G.T.G.A.A.A.G	Haplotypic	0.0002
PANSS Positive	C.T.T.A.T.A.T.C.G.T.G.A.G.G	Haplotypic	0.0337
PANSS Positive	rs10492354	Genotypic	0.0417
PANSS Positive		Allelic	0.0141
PANSS Negative	rs6427540	Genotypic	0.0251
PANSS Negative		Allelic	0.0380
PANSS Negative	G.C.T	Haplotypic	0.0380
PANSS Negative	G.C.C	Haplotypic	0.0046
PANSS Negative	rs6741819	Allelic	0.0361
PANSS Negative	T.A	Haplotypic	0.0469
PANSS Negative	rs230532	Genotypic	0.0130
PANSS Negative	rs230526	Genotypic	0.0245
PANSS Negative		Allelic	0.0232
PANSS Negative	rs118882	Genotypic	0.0192
PANSS Negative	rs230505	Genotypic	0.0101
PANSS Negative		Allelic	0.0025
PANSS Negative	rs230504	Genotypic	0.0001
PANSS Negative		Allelic	0.0121
PANSS Negative	rs230492	Genotypic	0.0246
PANSS Negative	rs230493	Genotypic	0.0000
PANSS Negative		Allelic	0.0069
PANSS Negative	rs230495	Genotypic	0.0016
PANSS Negative		Allelic	0.0007
PANSS Negative	rs3774959	Genotypic	0.0000
PANSS Negative	C.A.T.A.C.A.G.T.G.A.A.A.G	Haplotypic	0.0479
PANSS Negative	C.T.T.A.T.A.G.C.G.T.A.A.G.G	Haplotypic	0.0015
PANSS Negative	rs17440909	Genotypic	0.0068
PANSS Negative		Allelic	0.0084
PANSS Negative	rs17742120	Genotypic	0.0078
PANSS Negative		Allelic	0.0302
PANSS Negative	C.A	Haplotypic	0.0246
PANSS Negative	T.G	Haplotypic	0.0246
PANSS General	rs230532	Genotypic	0.0105
PANSS General	rs118882	Genotypic	0.0461
PANSS General	rs230504	Genotypic	0.0185
PANSS General		Allelic	0.0456
PANSS General	rs230493	Genotypic	0.0083
PANSS General		Allelic	0.0388

APPENDIX B

ASSOCIATION ANALYSES SUPPLEMENTARY DATA

Response measurement	Variant/ haplotype	Association	P-value
PANSS General	rs230495	Allelic	0.0312
PANSS General	rs3774959	Genotypic	0.0004
PANSS General		Allelic	0.0375
PANSS General	rs10492354	Genotypic	0.0458
PANSS General		Allelic	0.0164
PANSS Total	G.C.C	Haplotypic	0.0416
PANSS Total	rs230532	Genotypic	0.0083
PANSS Total	rs118882	Genotypic	0.0288
PANSS Total	rs230505	Allelic	0.0168
PANSS Total	rs230504	Genotypic	0.0028
PANSS Total		Allelic	0.0157
PANSS Total	rs230493	Genotypic	0.0007
PANSS Total		Allelic	0.0110
PANSS Total	rs230495	Genotypic	0.0138
PANSS Total		Allelic	0.0047
PANSS Total	rs3774959	Genotypic	0.0001
PANSS Total		Allelic	0.0133
PANSS Total	C.A.T.A.C.A.G.T.G.A.A.A.G	Haplotypic	0.0123
PANSS Total	rs10492354	Allelic	0.0361
Treatment-refractoriness	rs7520258	Genotypic	0.0380
Treatment-refractoriness	rs1352318	Allelic	0.0338
Remission	rs1352318	Genotypic	0.0083
Early treatment response	C.A.T.A.C.A.G.C.G.T.A.A.A.G	Haplotypic	0.0243
Weight	rs230532	Genotypic	0.0425
Weight	rs118882	Genotypic	0.0210
BMI	rs118882	Genotypic	0.0304
HDL	rs17727261	Allelic	0.0445
LDL	rs230534	Genotypic	0.0123
LDL	rs230532	Genotypic	0.0222
LDL	rs118882	Genotypic	0.0082
LDL	rs230520	Genotypic	0.0150
LDL	rs230504	Genotypic	0.0207
LDL		Allelic	0.0391
LDL	rs230492	Genotypic	0.0116
LDL		Allelic	0.0361
LDL	rs230539	Genotypic	0.0148
Total cholesterol	rs230534	Genotypic	0.0128
Total cholesterol	rs230532	Genotypic	0.0413
Total cholesterol	rs118882	Genotypic	0.0255
Total cholesterol	rs230520	Genotypic	0.0124
Total cholesterol	rs230492	Genotypic	0.0203
Total cholesterol	rs230539	Genotypic	0.0096
Total cholesterol	rs4648055	Genotypic	0.0459
Triglycerides	rs10218843	Genotypic	0.0398
Triglycerides		Allelic	0.0360

APPENDIX B

ASSOCIATION ANALYSES SUPPLEMENTARY DATA

Response measurement	Variant/ haplotype	Association	P-value
Triglycerides	A.T.C	Haplotypic	0.0492
Triglycerides	rs230532	Allelic	0.0237
Triglycerides	rs118882	Allelic	0.0209
Triglycerides	rs230504	Allelic	0.0288
Triglycerides	rs230493	Allelic	0.0271
Triglycerides	rs230539	Allelic	0.0412
Triglycerides	rs4648055	Allelic	0.0442
Triglycerides	C.A.T.A.C.A.G.C.G.T.A.A.A.G	Haplotypic	0.0237
Triglycerides	rs10492354	Allelic	0.0389

APPENDIX C: Conference outputs

Poster presentations

Ovenden, E.S., Drögemöller, B.I., Emsley, R.A., Warnich, L. 2015. Investigating the functional significance of genome-wide variants associated with treatment response in schizophrenia. The XXIIIrd World Congress of Psychiatric Genetics (WCPG). 16-20 October. Toronto, Canada.

Awarded Early Career Investigator Program (ECIP) Travel Award.

Ovenden, E.S., Drögemöller, B.I., Roetz, N.J., Emsley, R.A., Warnich, L. 2013. Investigating the functional significance of GWAS associations with antipsychotic treatment response in schizophrenia. The 15th Biennial Conference of the Southern African Society for Human Genetics (SASHG). 6-9 October. Johannesburg, South Africa.

Oral presentations

Ovenden, E.S., Drögemöller, B.I., Emsley, R.A., Warnich, L. 2015. Investigating the functional significance of genome-wide variants associated with treatment response in schizophrenia. Pharmacogenetics in Psychiatry (PIP). 15 October. Toronto, Canada.

Awarded Young Investigator Travel Award.

Ovenden, E.S., Drögemöller, B.I., Emsley, R.A., Warnich, L. 2015. Investigating the functional significance of genome-wide variants associated with treatment response in schizophrenia. The 16th Biennial Congress of the Southern African Society for Human Genetics (SASHG). 16-19 August. Pretoria, South Africa.

Ovenden, E.S., Drögemöller, B.I., Emsley, R.A., Warnich, L. 2014. Investigating the functional significance of genome-wide variants associated with antipsychotic treatment response. Joint South African Society for Bioinformatics and South African Genetics Society (SASBi-SAGS) Congress 2014. 23-26 September. Pretoria, South Africa.

Awarded best SASBi MSc oral presentation.

Ovenden, E.S., Drögemöller, B.I., Ishaque, N., Emsley, R.A., Warnich, L. 2014. Investigating the functional significance of genome-wide variants associated with antipsychotic treatment response. The 17th World Congress of Basic and Clinical Pharmacology (WCP). 13-18 July. Cape Town, South Africa.