

Investigation of genetic variation contributing to antipsychotic treatment response in a South African first episode schizophrenia cohort.

Britt Ingrid Drögemöller



Dissertation presented for the degree of Doctor of Philosophy (PhD) in Genetics at Stellenbosch University.

Supervisor:

Professor Louise Warnich

Co-supervisors:

Professor Robin Emsley

Professor Dana Niehaus

December 2013

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole 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.

This dissertation includes one original paper published in a peer-reviewed journal, one original paper that is in press in a peer-reviewed journal, one publication that is currently under peer-review and one unpublished publication. The development and writing of the papers (published and unpublished) were the principal responsibility of myself and, for each of the cases where this is not the case, a declaration is included in the dissertation indicating the nature and extent of the contributions of co-authors.

Date: December 2013

Copyright © 2013 Stellenbosch University of Stellenbosch

All rights reserved

Abstract

Schizophrenia is a debilitating disorder that occurs the world over. Although antipsychotics are largely effective in treating the positive symptoms of schizophrenia, the outcomes are non-optimal in many patients. As antipsychotic treatment response has been shown to be heritable, it is expected that the implementation of antipsychotic pharmacogenomics should aid in the optimization of antipsychotic treatments, however to date clinically applicable results are limited. Therefore this study utilized exome sequencing in a cohort of well characterized first episode schizophrenia patients to identify the genetic factors contributing to antipsychotic treatment response.

The utility of exome sequencing for antipsychotic pharmacogenomic applications in the African context was assessed through examination of the literature and publically available data. Thereafter, a cohort of 104 well characterized South African first episode schizophrenia patients who were treated with flupenthixol decanoate for twelve months was collected. From this cohort, subsets of patients on extreme ends of the treatment response spectrum were identified for exome sequencing. Thereafter a bioinformatics pipeline was used to call and annotate variants. These variants and those that have previously been associated with antipsychotic response, along with a panel of ancestry informative markers, were prioritized for genotyping in the entire cohort of patients. After genotyping of the 393 variants, statistical analyses were performed to identify associations with treatment response outcomes.

Examination of the literature revealed a need for exome sequencing in Africa. However, critical analyses of next generation sequencing data demonstrated that complex regions of the genome may not be well suited to these technologies. Thus, it may be necessary to combine exome sequencing with knowledge obtained from past research, as was done in this study to identify the genetic factors contributing to antipsychotic treatment response. Using this strategy, the current study highlighted the potential role that rare variants play in antipsychotic treatment response and additionally detected 11 variants that were significantly associated with antipsychotic treatment response outcomes ($P=2.19 \times 10^{-5}$). Nine of these variants were predicted to alter the function of the genes in which they occurred; of which eight were novel with regards to antipsychotic treatment response. The remaining two variants have been associated with antipsychotic treatment outcomes in previous GWAS. Examination of the function of the genes in which the variants occurred revealed that the variants associated with (i) positive symptom improvement were involved in the folate metabolism pathway and (ii) negative and general pathological symptoms improvement had potential links to neuronal development and migration.

To our knowledge this study is the first to utilize exome sequencing for antipsychotic pharmacogenomic purposes. The ability of this study to identify significant associations, even after correction for multiple testing, has highlighted the importance of combining genomic technologies with well characterized cohorts. The results generated from this study have served both to replicate results from previous antipsychotic pharmacogenetic studies and to identify novel genes and pathways involved in antipsychotic response. These results should aid in improving our understanding of the biological underpinnings of antipsychotic treatment response and may ultimately aid in the optimization of these treatments.

Opsomming

Skisofrenie is 'n siekte wat wêreldwyd voorkom en lei tot erge funksionele inkorting. Alhoewel antipsigotiese medikasie redelik effektief is in die behandeling van die positiewe simptome van skisofrenie, is die funksionele uitkomst in baie pasiënte nie optimaal nie. Die reaksie op antipsigotiese behandeling blyk oorerflik te wees. Die verwagting is dus dat die implementering van antipsigotiese farmakogenomika met die optimalisering van antipsigotiese behandeling sal help. Tot dusver het die resultate van farmakogenomika studies egter beperkte kliniese toepassings opgelewer. Hierdie studie het dus eksoom-volgordebepaling in 'n groep van goed-karakteriseerde eerste-episode skisofrenie pasiënte gebruik om die genetiese faktore wat bydra tot die antipsigotiese behandelings-reaksies te identifiseer.

Die gebruik van eksoom-volgordebepaling vir antipsigotiese farmakogenomika in die Afrika-konteks is deur die ondersoek van literatuur en openbaar-beskikbare data geëvalueer. Daarna is 'n groep van 104 goed-gekarakteriseerde Suid-Afrikaanse eerste-episode skisofrenie pasiënte, wat met flupenthixol dekanooat vir twaalf maande behandel is, versamel. Uit hierdie groep is subgroepe van pasiënte op die teenoorgestelde eindpunte van die behandelings-reaksiespektrum vir eksoom-volgordebepaling geïdentifiseer. Hierna is 'n bioinformatika pyplyn gebruik om variante te identifiseer en te annoteer. Hierdie variante, asook variante wat voorheen met antipsigotiese reaksie geassosieer is, is saam met 'n paneel van afkoms-informatiewe merkers vir genotipering in die hele groep pasiënte geprioritiseer vir genotipering. Na genotipering van die 393 variante, is statistiese analises uitgevoer om assosiasies met behandelings-reaksie uitkomst te identifiseer.

'n Ondersoek van die literatuur het getoon dat daar 'n behoefte vir eksoom-volgordebepaling in Afrika is. 'n Kritiese analise van volgende-generasie volgordebepalings data het egter getoon dat komplekse dele van die genoom nie geskik is vir die gebruik van hierdie tegnologie nie. Om die genetiese faktore wat bydra tot suksesvolle antipsigotiese behandeling te identifiseer, mag dit nodig wees om eksoom-volgordebepaling te kombineer met bevindings verkry uit vorige navorsing, soos gedoen in hierdie studie. In die huidige studie het die gebruik van hierdie strategie die potensiële rol van skaars variante in antipsigotiese behandelings-reaksies beklemtoon en 'n bykomende 11 variante is geïdentifiseer wat beduidend met antipsigotiese behandelingsrespons geassosieer is ($P=2.19 \times 10^{-5}$). Daar is voorspel dat nege van hierdie variante die funksie van die gene waarin hulle voorkom sal verander en agt van hierdie variante is vir die eerste keer met antipsigotiese behandelingsrespons geassosieer. Die oorblywende twee variante is met antipsigotiese behandelingsrespons in vorige GWAS geassosieer. 'n Ondersoek na die funksie van die gene waarin die variasies voorgekom het, toon dat die variante wat geassosieer is met (i) verbetering van positiewe simptome 'n rol speel in folaat-metabolisme, terwyl variante wat geassosieer is met (ii) die verbetering in negatiewe en algemene patologiese simptome potensiële skakels met neuron ontwikkeling en migrasie het.

Na ons wete is hierdie die eerste studie wat eksoom-volgordebepaling vir antipsigotiese farmakogenomika doeleindes gebruik. Die vermoë van hierdie studie om beduidende assosiasies te identifiseer, selfs na korreksie vir veelvoudige toetse, onderstreep die belangrikheid van die kombinerings van genomiese tegnologie met goed-gekarakteriseerde pasiënte. Die bevindinge van hierdie studie het nie net die resultate van vorige

antipsigotiese farmakogenetiese studies bevestig nie, maar ook nuwe gene en variante wat betrokke is in antipsigotiese reaksie geïdentifiseer. Hierdie resultate sal hopelik ons begrip van die onderliggende biologiese faktore wat antipsigotiese behandelingsrespons beïnvloed verbeter en uiteindelik ook met die optimalisering van behandeling help.

Acknowledgements

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

My supervisor, Prof Louise Warnich (Department of Genetics, Stellenbosch University), for mentoring, encouraging and inspiring me throughout my postgraduate studies, as well as for providing me with the opportunities to grow as a scientist.

My co-supervisors, Prof Robin Emsley and Prof Dana Niehaus (Department of Psychiatry, Stellenbosch University), for their constant support, guidance and motivation, as well as providing me with insight into the clinical aspects of this study.

Dr Galen Wright (Department of Genetics, Stellenbosch University and *the South African National Bioinformatics Institute, University of the Western Cape) for challenging and supporting me in all aspects, as well as for his novel ideas and computational assistance.
*current affiliation

Lundi Korkie (Department of Genetics, Stellenbosch University) for gDNA extractions, for ensuring that the lab runs smoothly and efficiently and for support and encouragement.

The rest of Lab 231 (Department of Genetics, Stellenbosch University) for providing constant support, encouragement and a happy work environment.

Dr Bonginkosi Chiliza, Dr Laila Asmal, Anneke Goosen, Patricia van Zyl, Talita Pienaar, Janette Jordaan, as well as the other members of the EONKCS team (Department of Psychiatry, Stellenbosch University) for providing me with assistance regarding the clinical aspects of the study.

Dr Anil Malhotra, Dr Todd Lencz (Zucker Hillside Hospital, NY, USA), Dr Tim Cutts and Dr Aarno Palotie (Sanger, UK) for providing advice regarding sequencing technologies.

HudsonAlpha genomics service lab (Alabama, USA) for exome sequencing.

Dr Steffen Durinck (Genetech, USA), Heine de Jager, Beatrix Coetzee, Paul Erasmus, Johan Loubser and Anelda van der Walt (from various departments and support services, Stellenbosch University) for providing me with guidance regarding the bioinformatic aspects of this project.

Prof Eileen Hoal and Michelle Daya (Division of Molecular Biology and Human Genetics, Stellenbosch University) for designing and analysing the AIMs data.

Dr Michael Klein (University of Utah Genomics Core Facility, USA) for the BeadXpress genotyping.

Prof Rouvay Roodt-Wilding, Dr Clint Rhode and Lise Sandenbergh (Department of Genetics, Stellenbosch University) for assistance with the GenomeStudio analyses.

Prof Lize van der Merwe (Department of Statistics, University of the Western Cape and Department of Molecular Biology and Human Genetics, Stellenbosch University) for assistance with the statistical analyses.

Anne Drögemöller for proof reading.

The Cold Spring Harbor conference organising committee, Biological Psychiatry conference organising committee, Southern African Society for Human Genetics conference organising committee, Oppenheimer Trust, South African National Research Foundation (NRF) and WhiteSci for providing travel bursaries to allow me to attend the various conferences.

The financial assistance of the South African NRF, Medical Research Council (MRC), Department of Science and Technology (DST)'s New Partnership for Africa's Development (NEPAD) grant and the L'Oréal-UNESCO for Women in Science in Sub-Saharan Africa fellowship towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily attributed to these organisations.

The results of this study have been presented at the following conferences:

Drögemöller B, Wright G, Niehaus D, Emsley R, Warnich L. The sequencing era: Are pharmacogenetic profiles becoming a reality? Pharmacogenomics and Personalized Therapy Meeting. 17-21 November 2010. Cold Spring Harbor Laboratory, USA. Poster presentation.

Drögemöller B, Wright G, Niehaus D, Emsley R, Warnich L. The sequencing era: Are pharmacogenetic profiles becoming a reality? Joint International Conference of the African and Southern African Societies of Human Genetics. 6 – 9 March 2011. Cape Town, South Africa. Poster presentation.

Drögemöller B, Niehaus D, Wright G, Chiliza B, Koen L, Emsley R, Warnich L. Personalised genomic profiles: applications in African psychiatric pharmacogenetics. Biological Psychiatry Congress. 22-25 September 2011. Stellenbosch, South Africa. Oral presentation.

– *Awarded best oral presentation by a young investigator.*

Drögemöller B, Niehaus D, Wright G, Chiliza B, Asmal L, Koen L, Emsley R, Warnich L. Mining schizophrenia exome data in search of variants contributing to antipsychotic treatment response. South African Genetics & Bioinformatics Society Conference. 10 –12 September 2012. Stellenbosch, South Africa. Oral presentation.

Drögemöller B, Niehaus D, Wright G, Chiliza B, Asmal L, Koen L, Emsley R, Warnich L. Well characterised cohorts and exome data: investigating antipsychotic pharmacogenomics in the South African context. The American Society for Human Genetics Annual Meeting. 6-10 November 2012. San Francisco, USA. Poster presentation.

Drögemöller B, Niehaus D, Wright G, Chiliza B, Asmal L, Koen L, Emsley R, Warnich L. Well characterised cohorts and exome data: investigating antipsychotic pharmacogenomics in the South African context. Personal Genomes and Medical Genomics Meeting. 14-17 November 2012. Cold Spring Harbor Laboratory, USA. Poster presentation.

Drögemöller B, Niehaus D, Wright G, Chiliza B, Asmal L, Emsley R, Warnich L. The use of exome sequencing for antipsychotic pharmacogenomic application in South African schizophrenia patients. African Society of Human Genetics Meeting. 19 – 21 May, 2013. Accra, Ghana. Poster presentation.

– *Awarded best poster presentation by a young investigator.*

Drögemöller B, Niehaus D, Wright G, Chiliza B, Asmal L, Emsley R, Warnich L. The use of exome sequencing for antipsychotic pharmacogenomic application in South African schizophrenia patients. Biological Psychiatry Congress. 29 August – 1 September, 2013. Wild Coast Sun, South Africa. Poster presentation.

Drögemöller B, Niehaus D, Chiliza B, van der Merwe L, Wright G, Daya M, Hoal E, Asmal L, Emsley R, Warnich L. The identification of novel variants associated with antipsychotic treatment response with the use of exome sequencing. Southern African Society for Human Genetics Congress. To be presented 6 – 9 October, 2013. Johannesburg, South Africa. Oral presentation.

The results of this study have generated the following publications:

Drögemöller BJ, Wright GEB, Niehaus DJH, Emsley R, Warnich L. Whole genome resequencing in pharmacogenomics: moving away from past disparities to globally representative applications. *Pharmacogenomics*. 12, 1717-28 (2011).

- A version of this paper is reprinted in Chapter 3, with permission from Future Medicine and is available at <http://www.futuremedicine.com/toc/pgs/12/12>

Drögemöller BJ, Wright GEB, Niehaus DJH, Emsley R, Warnich L. Next Generation Sequencing of Pharmacogenes: A Critical Analysis Focusing on Schizophrenia Treatment. *Pharmacogenetics and Genomics*. doi: 10.1097/FPC.0000000000000006 (2013).

- A version of this paper is reprinted in Chapter 4, with permission from Wolters Kluwer Health and is available at <http://journals.lww.com/jpharmacogenetics/toc/publishahead>

Drögemöller BJ, Niehaus DJH, Chiliza B, Asmal L, Malhotra AK, Wright GEB, Emsley R, Warnich L. Patterns of Variation Influencing Antipsychotic Treatment Outcomes in South African First Episode Schizophrenia Patients. *Pharmacogenomics*. In press.

- A version of this paper is reprinted in Chapter 5, with permission from Future Medicine.

Drögemöller BJ, Chiliza B, van der Merwe L, Wright GEB, Daya M, Hoal E, Asmal L, Niehaus DJH, Emsley R, Warnich L. The identification of novel genetic variants and pathways affecting antipsychotic treatment response in South African first episode schizophrenia patients. In preparation.

- A version of this paper is reprinted in Chapter 6.

Other publications related to this work can be found at:

<http://bit.ly/BDrogemollerPublications>

Dedications

This thesis is dedicated to my parents, Hugo and Anne Drögemöller, for their contribution to my genome and for providing an environment that encouraged a love for science and creativity.

Table of contents

List of figures.....	xii
List of supplementary figures	xii
List of tables	xii
List of supplementary tables.....	xiii
List of symbols and abbreviations.....	xiv
CHAPTER 1: Introduction	1
CHAPTER 2: Literature review.....	5
2.1 Schizophrenia.....	5
2.1.1 Clinical aspects of schizophrenia	5
2.1.2 Risk factors for schizophrenia.....	8
2.1.3 The genetics of schizophrenia	8
2.2 The antipsychotic treatment of schizophrenia	13
2.2.1 Background.....	13
2.2.2 Response to antipsychotic treatment	16
2.2.3 Adverse drug reactions caused by antipsychotic treatment.....	20
2.3 Antipsychotic pharmacogenomics.....	22
2.3.1 Background.....	22
2.3.2 Candidate gene antipsychotic pharmacogenetic studies.....	25
2.3.3 Genome-wide association studies for antipsychotic pharmacogenomics	30
2.4 The study of antipsychotic pharmacogenomics in the South African context	33
2.4.1 Resource limited settings and unique cultural considerations	33
2.4.2 The rainbow nation and implications for genomic studies	35
2.5 Aims and objectives	37
2.6 Study outline	38
PART 1: Evaluation of the utility of whole genome and exome sequencing for antipsychotic pharmacogenomics in Africa	39
CHAPTER 3: Whole genome resequencing in pharmacogenomics: moving away from past disparities to globally representative applications	40
3.1 Summary	40
3.2 Inequalities in genomic research	40
3.3 Next generation sequencing technologies	42
3.4 Pharmacogenomics: A front runner for translation of genomic results into the clinical setting?	43
3.5 Critical evaluation of the variation present in pharmacogenes	44
3.6 Hurdles associated with sequencing technologies	48
3.7 Conclusions	50
3.8 Future perspectives	52
CHAPTER 4: Next generation sequencing of pharmacogenes: a critical analysis focusing on schizophrenia treatment	54
4.1 Summary	54
4.2 Introduction	54
4.3 Materials and methods.....	56
4.3.1 Identification of candidate antipsychotic pharmacogenes and very important pharmacogenes	56

4.3.2	Critical analyses of factors potentially influencing the sequencing of pharmacogenes.....	57
4.3.3	Assessment of pharmacogenes using the 1000 Genomes Project mask files.....	57
4.3.4	Statistical analyses.....	57
4.4	Results.....	58
4.5	Discussion.....	61
4.6	Conclusion.....	63
	Summary of Chapters 3 and 4.....	65
	PART 2: Utilisation of exome sequencing for antipsychotic pharmacogenomics in the South African context.....	66
	CHAPTER 5: Patterns of variation influencing antipsychotic treatment outcomes in South African first episode schizophrenia patients	67
5.1	Summary	67
5.2	Introduction	67
5.3	Materials and methods.....	68
5.3.1	Patient samples	68
5.3.2	Exome sequencing.....	69
5.3.3	Bioinformatics pipeline for exome sequence analysis	69
5.3.4	Comparison of the patterns of variation observed in the non-responder and responder individuals.....	69
5.4	Results.....	70
5.5	Discussion.....	72
5.6	Conclusions	74
5.7	Future perspectives	74
	CHAPTER 6: The identification of novel genetic variants and pathways affecting antipsychotic treatment response in South African first episode schizophrenia patients	76
6.1	Summary	76
6.2	Introduction	77
6.3	Materials and Methods.....	78
6.3.1	Patient samples	78
6.3.2	Prioritisation of variants for genotyping in the entire FES cohort	79
6.3.3	Genotyping of the prioritised variants in the entire FES cohort	80
6.3.4	Statistical analyses.....	80
6.4	Results.....	81
6.5	Discussion.....	88
6.6	Conclusions	92
	Summary of Chapters 5 and 6.....	95
	CHAPTER 7: Conclusions, study limitations and future directions	96
7.1	Conclusions	96
7.1.1	PART 1: Evaluation of the utility of exome sequencing for antipsychotic pharmacogenomics in the African context	96
7.1.2	PART 2: The utilisation of exome sequencing for antipsychotic pharmacogenomics in the South African context.....	97
7.2	Study limitations	100
7.3	Future directions.....	102

APPENDIX 1: Analyses performed in GALAXY	105
APPENDIX 2: Next generation sequencing analyses of antipsychotic pharmacogenes.....	107
APPENDIX 3: Specified protocols, reagents and solutions.....	117
Specified protocols.....	117
Reagents and solutions.....	118
APPENDIX 4: Exome data generation	120
Preparation of gDNA samples.....	120
Quality control of gDNA samples.....	120
Quality control of sequence reads.....	121
APPENDIX 5: Variant calling and annotation of exome data	125
Transfer of data.....	125
Submission of jobs to the high performance computer	125
The Genome Analysis Toolkit (GATK): variant calling	125
APPENDIX 6: Comparison of the non-responder and responder exome data	130
APPENDIX 7: Selection of variants to genotype in the larger cohorts.....	133
Variants previously associated with antipsychotic response	133
Previous associations obtained from genome-wide association studies	133
Previous associations detected in candidate genes	133
Variants identified from the exome data	134
Functional variants detected in candidate genes	134
Variant Annotation, Analysis and Search Tool (VAAST) gene-based analyses	135
Variant Annotation, Analysis and Search Tool (VAAST) variant-based analyses	140
APPENDIX 8: Genotyping protocols for prioritised variants	142
Polymerase chain reaction amplification.....	142
Sanger sequence verification.....	142
Restriction fragment length polymorphism analyses.....	143
TaqMan® SNP genotyping.....	144
Illumina BeadXpress Assay genotyping.....	144
APPENDIX 9: Genotyping results for prioritised variants.....	146
PCR-RFLP and TaqMan Genotyping	146
Illumina BeadXpress genotyping results.....	147
APPENDIX 10: Statistical analyses.....	150
References	154

List of figures

Figure 1.1: Disparities in mental health care services.	2
Figure 2.1: A representation of the overlap in symptoms observed between schizophrenia and other psychiatric disorders.....	7
Figure 2.2: The progression of studies examining the genetic factors contributing to schizophrenia susceptibility.....	9
Figure 2.3: Relative contribution of common and rare variants to the development of schizophrenia.....	10
Figure 2.4: International Psychopharmacology Algorithm Project Dosing Algorithm.....	14
Figure 2.5: The central features of response to antipsychotic treatment.....	17
Figure 2.6: The role that treatment interventions have on reducing the burden of schizophrenia.....	22
Figure 2.7: The role that pharmacogenomics can play in improving antipsychotic treatment outcomes.	23
Figure 2.8: The differences in the frequencies of variants contributing to <i>CYP2D6</i> metabolising function across the globe.	26
Figure 2.9: The suggested dosage adjustments for various antidepressants and antipsychotics based on <i>CYP2D6</i> metaboliser status.....	26
Figure 3.1 Percentage African “GWAS hits”.....	41
Figure 3.2 Differences in pharmacogenetic allele frequencies.....	47
Figure 3.3: A comparison of the number of individuals for which genotype information was not available for the examined SNPs, stratified according to the date that the sequencing data was published..	49
Figure 4.1: <i>DRD4</i> and the corresponding inaccessible regions of the gene..	58
Figure 4.2: The percentage sequence similarity between <i>CYP2D6</i> and the corresponding pseudogenes as calculated by mVISTA.....	62
Figure 5.1 Comparison of the coding variation present in the non-responder and responder individuals.....	71
Figure 5.2: The percentage novel variation per individual observed for each class of variant.....	71
Figure 6.1: The selection of variants for genotyping using the BeadXpress Assay.....	80
Figure 6.2: A mosaic plot representing the observed genotype frequencies for rs2027937	82
Figure 6.3: The differences in the patterns of PANSS scores for the variants with the most significant <i>P</i> values for each of the symptom domains.	85
Figure 6.4: Proportions of ancestry, as determined by ADMIXTURE analyses of the 95 AIMs, observed in each of the SAC FES individuals that were examined.....	86
Figure 6.5: Minor allele frequency comparisons of the eleven polymorphisms that were significantly associated with antipsychotic pharmacogenomics phenotypes in the current study	87

List of supplementary figures

Figure S 1: The integrity of gDNA samples LW0001-LW0011	120
Figure S 2: Alignment and sorting of generated reads	122
Figure S 3: Example of A: Observed fragment sizes; B: Observed quality scores; C: Observed bases per cycle; D: Observed quality scores per cycle.	123
Figure S 4: Variant calling pipeline in the GATK.....	129
Figure S 5: The models of disease that were used in the VAAST analyses	135
Figure S 6: TaqMan SNP genotyping results for rs11368509 in <i>UPP2</i>	147
Figure S 7: BeadXpress genotyping clusters visualised in GenomeStudio.....	149
Figure S 8: Observed trait values represented as box plots and corresponding percentage change from baseline in trait values.	153

List of tables

Table 1.1: Warfarin dosing calculator.....	3
Table 2.1: Pharmacogenetic tests that have been designed for antipsychotic treatment.....	24
Table 2.2: The top 25 most frequently studied candidate antipsychotic genes.....	28
Table 2.3: The GWAS and top variants significantly associated with antipsychotic pharmacogenetic traits as reported on the HuGE Navigator GWAS Integrator.....	31
Table 3.1: Novel/uncharacterised non-synonymous SNPs predicted to be damaging, from resequencing data of 10 pharmacogenes in 13 ethnically diverse individuals.....	45
Table 3.2: List of important previously characterised variants affecting treatment response.....	46
Table 3.3: Allele combination comparisons utilising data from our analyses to compare to that of Ng <i>et al.</i> [319].....	50
Table 4.1: The genomic composition of antipsychotic pharmacogenes that are also considered to be very important pharmacogenes.....	59
Table 4.2: The genomic composition of those genes with more than 50% “inaccessible genome”	59
Table 5.1: The total and average number of variants observed in the five non-responder and responder individuals.....	70
Table 5.2: The allele frequencies of the loss of function variants present in the non-responder siblings, but absent from the responder individuals.....	72
Table 6.1: PANSS score assessments at baseline and twelve months.....	79
Table 6.2: The effect sizes and corresponding 95% confidence intervals of the significantly associated variants.....	83
Table 6.3: The significantly associated variants and their predicted effects on traits.....	84
Table 6.4: Maximum, minimum and median proportions of ancestry observed in the SAC FES individuals.....	86

List of supplementary tables

Table S 1: Thirteen ethnically diverse individuals utilised to compare pharmacogene variation.....	106
Table S 2: The percentage sequence similarity, GC content, low complexity sequence and “inaccessible genome” associated with each of the pharmacogenes.....	109
Table S 3: Quality assessments for the gDNA samples.....	121
Table S 4: Summary of sequence coverage obtained for the targeted exomes of the eleven FES samples.....	122
Table S 5: Sequence reads generated and aligned to the hg19 reference sequence.....	124
Table S 6: The GATK resources.....	126
Table S 7: Primer sequences for amplification of regions containing the variants that were not genotyped by the Illumina BeadXpress Assay.....	143
Table S 8: RFLP specifications.....	145
Table S 9: Sanger sequencing conformation and PCR-RFLP genotyping of variants that were not genotyped with the Illumina BeadXpress Assay.....	146
Table S 10: SNPs that failed the Illumina BeadXpress genotyping.....	148

List of symbols and abbreviations

3'	3-prime end
5'	5-prime end
©	Copyright
°C	Degrees celsius
\$	Dollar
=	Equal to
>	Greater than
µg	Microgram
µl	Microlitre
µM	Micromolar
%	Percentage
p	Pico
®	Registered trademark
<	Smaller than
X	Times
™	Trademark
A	Adenine
A	Alanine
<i>ABCB1</i>	ATP-binding cassette, sub-family B (MDR/TAP), member 1 gene
<i>ADO</i>	2-aminoethanethiol (cysteamine) dioxygenase gene
<i>ADRB1</i>	Adrenoceptor beta 1 gene
<i>ADRB2</i>	Adrenoceptor beta 2 gene
ADRs	Adverse drug reactions
ADT	Assay design tool
AFR	African 1000 genomes project population
AIDS	Acquired immunodeficiency syndrome
AIMs	Ancestry informative markers
<i>ANKRD53</i>	Ankyrin repeat domain 53 gene
<i>ANKS1B</i>	Ankyrin repeat and sterile alpha motif domain containing 1B gene
ASN	Asian 1000 genomes project population
ASW	African ancestry in Southwest United States of America
<i>BDNF</i>	Brain-derived neurotrophic factor gene
BLAST	Basic local alignment search tool
Bp	Base pair
BPRS	Brief psychiatric rating scale
<i>BRCA1</i>	Breast cancer 1, early onset gene
BSA	Bovine serum albumin
BWA	Burrows-Wheeler alignment
C	Cytosine
<i>C6</i>	Complement component 6 gene
<i>C6orf52</i>	Chromosome 6 open reading frame 52 gene
CAFE	Comparisons of atypicals in first episode
CATIE	Clinical antipsychotic trials of intervention effectiveness
<i>CCHCR1</i>	Coiled-coil alpha-helical rod protein 1 gene
<i>CERKL</i>	Ceramide kinase-like gene
CEU	Utah residents with Northern and Western European ancestry from the CEPH collection

<i>CHFR</i>	Checkpoint with forkhead and ring finger domains, E3 ubiquitin protein ligase gene
ChIP-seq	Chromatin immunoprecipitation sequencing
CI	Confidence interval
<i>CLLU10S</i>	Chronic lymphocytic leukemia up-regulated 1 opposite strand gene
CNVs	Copy number variants
<i>COMT</i>	Catechol-O-methyltransferase gene
CPIC	Clinical pharmacogenetics implementation consortium
CUtLASS	Cost utility of the latest antipsychotic drugs in schizophrenia
<i>CYP</i>	Cytochrome P450 gene
<i>CYP1A2</i>	Cytochrome P450, family 1, subfamily A, polypeptide 2 gene
<i>CYP2A6</i>	Cytochrome P450, family 2, subfamily A, polypeptide 6 gene
<i>CYP2B6</i>	Cytochrome P450, family 2, subfamily B, polypeptide 6 gene
<i>CYP2C9</i>	Cytochrome P450, family 2, subfamily C, polypeptide 9 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
<i>CYP4F2</i>	Cytochrome P450, family 4, subfamily F, polypeptide 2 gene
<i>CYTH1</i>	Cytohesin 1 gene
D	Aspartic acid
DALYs	Disability adjusted life years
<i>DISC1</i>	Disrupted in schizophrenia 1 gene
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
<i>DNAAF1</i>	Dynein, axonemal, assembly factor 1 gene
DNase-seq	DNase I hypersensitive sites sequencing
<i>DNHD1</i>	Dynein heavy chain domain 1 gene
<i>DNMBP</i>	Dynamin binding protein gene
dNTPs	Deoxynucleotide triphosphates
<i>DRD1</i>	Dopamine receptor D1 gene
<i>DRD2</i>	Dopamine receptor D2 gene
<i>DRD3</i>	Dopamine receptor D3 gene
<i>DRD4</i>	Dopamine receptor D4 gene
<i>DRD5</i>	Dopamine receptor D5 gene
DSM-IV	Diagnostic and statistical manual of mental disorders version four
DSM-5	Diagnostic and statistical manual of mental disorders version five
DST	Department of science and technology
<i>DTNBP1</i>	Dystrobrevin binding protein 1 gene
DUP	Duration of untreated psychosis
E	Glutamic acid
EDTA	Ethylenediaminetetraacetic Acid
e.g.	<i>Exempli gratia</i>
EM	Extensive metaboliser
EPS	Extrapyramidal symptoms
<i>et al.</i>	<i>Et alii</i>
EUFEST	European first episode schizophrenia trial
EUR	European 1000 genomes project population

F	Forward primer
F	Phenylalanine
FDA	Food and drug administration
FES	First episode schizophrenia
FGAs	First generation antipsychotics
<i>FIGN</i>	Fidgetin gene
<i>FREM3</i>	FRAS1 related extracellular matrix 3 gene
G	Gram
G	Guanine
<i>GABA</i>	Gamma-aminobutyric acid gene
<i>GATK</i>	Genome analysis toolkit
gDNA	Genomic deoxyribonucleic acid
<i>GLI2</i>	GLI family zinc finger 2 gene
<i>GLIS1</i>	GLIS family zinc finger 1 gene
<i>GNB3</i>	Guanine nucleotide binding protein (G protein), beta polypeptide 3 gene
<i>GPRIN2</i>	G protein regulated inducer of neurite outgrowth 2 gene
<i>GRIN3B</i>	Glutamate receptor, ionotropic, N-methyl-D-aspartate 3B gene
<i>GSTM1</i>	Glutathione S-transferase mu 1 gene
<i>GSTP1</i>	Glutathione S-transferase pi 1 gene
<i>GSTT1</i>	Glutathione S-transferase theta 1 gene
GWAS	Genome-wide associations studies
H	Histidine
H3Africa	Human, hereditary and health in Africa
HGNC	Human genome nomenclature committee
HGP	Human genome project
hiPSC	Human induced pluripotent cells
<i>HIST1H2BM</i>	Histone cluster 1, H2bm gene
HIV	Human immunodeficiency virus
<i>HK1</i>	Hexokinase 1 gene
<i>HLA-B</i>	Major histocompatibility complex, class I, B gene
<i>HLA-C</i>	Major histocompatibility complex, class I, C gene
<i>HLA-DQA1</i>	Major histocompatibility complex, class II, DQ alpha 1 gene
<i>HLA-DQB1</i>	Major histocompatibility complex, class II, DQ beta 1 gene
<i>HLA-DQB3</i>	Major histocompatibility complex, class II, DQ beta 3 gene
<i>HLA-DRB1</i>	Major histocompatibility complex, class II, DR beta 1 gene
<i>HLA-DRB5</i>	Major histocompatibility complex, class II, DR beta 5 gene
<i>HOMER</i>	Homeobox and leucine zipper encoding gene
HPC	High performance computer
hr	Hour
<i>HSPA1A</i>	Heat shock 70kDa protein 1A gene
<i>HTR1A</i>	5-hydroxytryptamine (serotonin) receptor 1A, G protein-coupled gene
<i>HTR2A</i>	5-hydroxytryptamine (serotonin) receptor 2A, G protein-coupled gene
<i>HTR2C</i>	5-hydroxytryptamine (serotonin) receptor 2C, G protein-coupled gene
<i>HTR6</i>	5-hydroxytryptamine (serotonin) receptor 6, G protein-coupled gene
HWE	Hardy-Weinberg equilibrium
I	Isoleucine
ICD-10	International classification of diseases version 10
ID	Identification

IM	Intermediate metaboliser
Inc.	Incorporation
<i>INSIG2</i>	Insulin induced gene 2 gene
IPAP	International psychopharmacology algorithm project
K	Lysine
Kb	Kilobase
<i>KRT5</i>	Keratin 5 gene
L	Leucine
L	Litre
LAI	Long-acting injectables
LD	Linkage disequilibrium
<i>LEP</i>	Leptin gene
<i>LEPR</i>	Leptin receptor gene
LMIC	Low- and middle income countries
LOD	Logarithm of odds
LOF	Loss of function
Ltd	Limited
LWK	Luhya in Webuye, Kenya
M	Methionine
M	Molar
m	Mutagenic primer
MAF	Minor allele frequency
MATRICES	Measurement and treatment research to improve cognition in schizophrenia
Mb	Megabases
<i>MC4R</i>	Melanocortin 4 receptor gene
Meth-seq	Methylation sequencing
Mg	Milligrams
MgCl ₂	Magnesium chloride
MHC	Major histocompatibility complex
Min	Minutes
<i>MIR137</i>	MicroRNA 137 gene
miRNA	Micro ribonucleic acid
MKK	Maasai in Kinyawa, Kenya
ml	Millilitre
mM	Millimolar
<i>MMP16</i>	Matrix metalloproteinase 16 (membrane-inserted) gene
MRC	Medical research council
<i>MTHFR</i>	Methylenetetrahydrofolate reductase (NAD(P)H) gene
<i>MTRR</i>	5-methyltetrahydrofolate-homocysteine methyltransferase reductase gene
<i>MUC16</i>	Mucin 16, cell surface associated gene
<i>MYL5</i>	Myosin, light chain 5, regulatory gene
n	Sample size
<i>N4BP3</i>	NEDD4 binding protein 3 gene
NA	Not applicable
NCBI	National centre for biotechnology information
ncRNA	Non-coding RNA
<i>NDOR1</i>	NADPH dependent diflavin oxidoreductase 1 gene

NEPAD	New partnership for Africa's development
Ng	Nanogram
NGS	Next generation sequencing
NIMH	National institute of mental health
nm	Nanometre
<i>NMDA</i>	N-methyl D-aspartate gene
<i>NOV</i>	Nephroblastoma overexpressed gene
<i>NPAS3</i>	Neuronal PAS domain protein 3 gene
<i>NQO1</i>	NAD(P)H dehydrogenase, quinone 1 gene
<i>NRAP</i>	Nebulin-related anchoring protein gene
NRF	National research foundation
OR	Odds ratio
<i>OR5AC2</i>	Olfactory receptor, family 5, subfamily AC, member 2 gene
<i>OXT</i>	Oxytocin/neurophysin I prepropeptide gene
p	Page
<i>P</i>	Probability
P	Proline
PANSS	Positive and negative syndrome scale
PANSS-G	Positive and negative syndrome scale-general pathological
PANSS-N	Positive and negative syndrome scale-negative
PANSS-P	Positive and negative syndrome scale-positive
PANSS-T	Positive and negative syndrome scale-total
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PGC	Psychiatric genome consortium
PharmGKB	The pharmacogenomics knowledge base
<i>PKD1L3</i>	Polycystic kidney disease 1-like 3 gene
<i>PLEKHG3</i>	Pleckstrin homology domain containing, family G (with RhoGef domain) member 3 gene
PM	Poor metabolisers
<i>PMCH</i>	Pro-melanin-concentrating hormone gene
PolyPhen	Polymorphism phenotyping
Prof	Professor
Pty	Proprietary limited company
Q	Glutamine
QC	Quality control
R	Arginine
R	Reverse primer
RFLP	Restriction fragment length polymorphism
<i>RGS4</i>	Regulator of G-protein signaling 4 gene
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
rpm	Revolutions per minute
rs	RefSNP

S	Serine
SAC	South African Coloured
SANS	Scale for the assessment of negative symptoms
SAPS	Scale for the assessment of positive symptoms
SDS	Sodium dodecyl sulfate
Sec	Seconds
SGAs	Second generation antipsychotics
SIFT	Sorting intolerant from tolerant
<i>SIPA1L2</i>	Signal-induced proliferation-associated 1 like 2 gene
SJS	Steven Johnson syndrome
<i>SLC6A3</i>	Solute carrier family 6 (neurotransmitter transporter, dopamine), member 3 gene
<i>SLC6A4</i>	Solute carrier family 6 (neurotransmitter transporter, serotonin), member 4 gene
<i>SLC22A23</i>	Solute carrier family 22, member 23 gene
<i>SLC25A21</i>	Solute carrier family 25 (mitochondrial oxoadipate carrier), member 21 gene
<i>SLC25A41</i>	Solute carrier family 25, member 41 gene
<i>SLC26A9</i>	Solute carrier family 26, member 9 gene
<i>SLC39A4</i>	Solute carrier family 39 (zinc transporter), member 4 gene
<i>SLCO3A1</i>	Solute carrier organic anion transporter family, member 3A1 gene
<i>SNAP25</i>	Synaptosomal-associated protein, 25kDa gene
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
<i>SOD2</i>	Superoxide dismutase 2, mitochondrial gene
SOHO	Schizophrenia outpatients health outcome study
<i>SPAG17</i>	Sperm associated antigen 17 gene
Ss	Submitter SNP
<i>STK33</i>	Serine/threonine kinase 33 gene
<i>SULT1A1</i>	Sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1 gene
T	Threonine
T	Thymine
<i>Taq</i>	<i>Thermus aquaticus</i>
TB	Tuberculosis
TBE	Tris borate ethylenediaminetetraacetic acid buffer
<i>TCF4</i>	Transcription factor 4 gene
<i>TCF25</i>	Transcription factor 25 gene
TD	Tardive dyskinesia
TE	Tris ethylenediaminetetraacetic acid buffer
TFPGA	Tools for population genetic analysis
<i>TGFB1</i>	Transforming growth factor, beta 1 gene
<i>TMEM235</i>	Transmembrane protein 235 gene
<i>TNF</i>	Tumor necrosis factor gene
<i>TPMT</i>	Thiopurine S-methyltransferase gene
<i>TRIM17</i>	Tripartite motif containing 17 gene
<i>TXNDC16</i>	Thioredoxin domain containing 16 gene
<i>TYMS</i>	Thymidylate synthetase gene
U	Unit (enzyme quantity)
UCSC	University of California Santa Cruz
<i>UGT1A1</i>	UDP glucuronosyltransferase 1 family, polypeptide A1 gene
UK	United Kingdom
UM	Ultra-rapid metabolisers

UNESCO	United nations educational, scientific and cultural organisation
<i>UPP2</i>	Uridine phosphorylase 2 gene
US	United States
USA	United States of America
UTR	Untranslated region
UV	Ultraviolet
V	Valine
V	Version
V	Volts
VAAST	Variant annotation, analysis and search tool
VIP	Very important pharmacogenes
<i>VKORC1</i>	Vitamin K epoxide reductase complex, subunit 1 gene
Vs	Versus
v/v	Volume per volume
WGES	Whole genome and exome sequencing
WHO	World health organisation
w/v	Weight per volume
<i>WWOX</i>	WW domain containing oxidoreductase gene
www	World wide web
Y	Tyrosine
YRI	Yoruban in Ibadan, Nigeria
<i>ZBTB24</i>	Zinc finger and BTB domain containing 24 gene
<i>ZC3H3</i>	Zinc finger CCCH-type containing 3 gene
<i>ZNF470</i>	Zinc finger protein 470 gene

CHAPTER 1: Introduction

More than 13% of the global burden of disease can be attributed to neuropsychiatric disorders [1] with approximately \$2.5 trillion spent per year in the USA on mental disorders alone [2]. This burden is, however, not equally shared and it has been reported that only 25% of the global burdens attributed to neuropsychiatric disorders fall on high income countries, with the already resource limited low- and middle income countries (LMIC) absorbing the rest of the burden [1]. This burden is heightened by inadequate treatments and it has been estimated that, in LMIC, only between 16-24% of people suffering from serious mental disorders receive treatment [3]. This may in part be attributed to a lack of resources, which are a result of inadequate funds and the high rate of other disease such as HIV/AIDS, with 70% of African countries spending less than 1% of their already small health care budgets on mental health [4]. The lack of mental health services in LMIC is further highlighted by the fact that there are approximately 200 fold less psychiatrists per 100 000 individuals in low income countries when compared to high income countries, with these disparities extending to other aspects of mental health care [2] (Figure 1.1). In conjunction with this, research disparities in LMIC are of serious concern. This is reflected by the 10/90 gap, which refers to the fact that only 10% of health research funding is allocated to health issues that account for 90% of the global disease burden [5]. Taking this a level further, there is a 5/95 gap with regards to publications, with only 5% of the research from LMIC published in high impact psychiatric journals [6]. This lack of research combined with the inability of the over-burdened health care systems to recognise mental health issues as serious health concerns, adds to the stigma and costs associated with these disorders [7]. As primary health care options are not available to psychiatric patients and communities are often not equipped to understand the biological underpinnings of these disorders, many individuals and their caregivers are placed under immense socio-economic burdens. For example, in Somalia, due to the lack of understanding and treatment of psychiatric disorders, family members pay for the “hyena cure”, which entails placing the individual suffering from a psychiatric illness into a pit of starved hyenas so as to rid the patient of *djinns* (evil spirits) [8].

One of the mental disorders most affected by these issues is schizophrenia, which is extremely complex and poorly understood [9,10]. Furthermore, the debilitating symptoms associated with this disorder make it both highly stigmatised and costly to treat [9]. The high costs associated with schizophrenia can be divided into direct and indirect costs, with direct costs relating to the costs of treating the disease (e.g. hospitalisations, health care costs and medications) and indirect costs relating to the burden of the disease (e.g. loss of productivity as a result of the disease) [11]. In high income countries, the direct and indirect costs related to schizophrenia are approximately equal, with most of the direct costs attributed to hospital admissions and only a small portion of expenses associated with the cost of medication. In LMIC on the other hand, the economic burden of schizophrenia is skewed towards indirect costs, with nearly 10 times as many Disability Adjusted Life Years (DALYs) attributed to schizophrenia in LMIC when compared to high income countries (15.2 million DALYs vs. 1.6 million DALYs) [11,12]. The unequal burden of the indirect costs associated with schizophrenia in LMIC can be attributed to the current lack of health care resources allocated to the treatment of this stigmatised and debilitating disorder, which is

largely overshadowed by the HIV/AIDS and TB epidemics [13]. Thus, very few schizophrenia patients will be hospitalised, which is reflected by the fact that in Nigeria, the cost of medication accounts for 61.8% of the direct costs associated with schizophrenia, in comparison to the 1.1-9.0% of direct costs in high income countries [11]. This may create the false impression that patients frequently receive medication in LMIC. However, as mentioned previously, 76-84% of patients with serious mental disorders living in these regions do not receive treatment [3]. Thus, the treatment of schizophrenia in LMIC requires urgent attention, which includes an increase in the resources allocated to health care and research. Although this will result in an initial increase in the direct costs associated with schizophrenia, the long term decrease in indirect costs should outweigh these concerns.

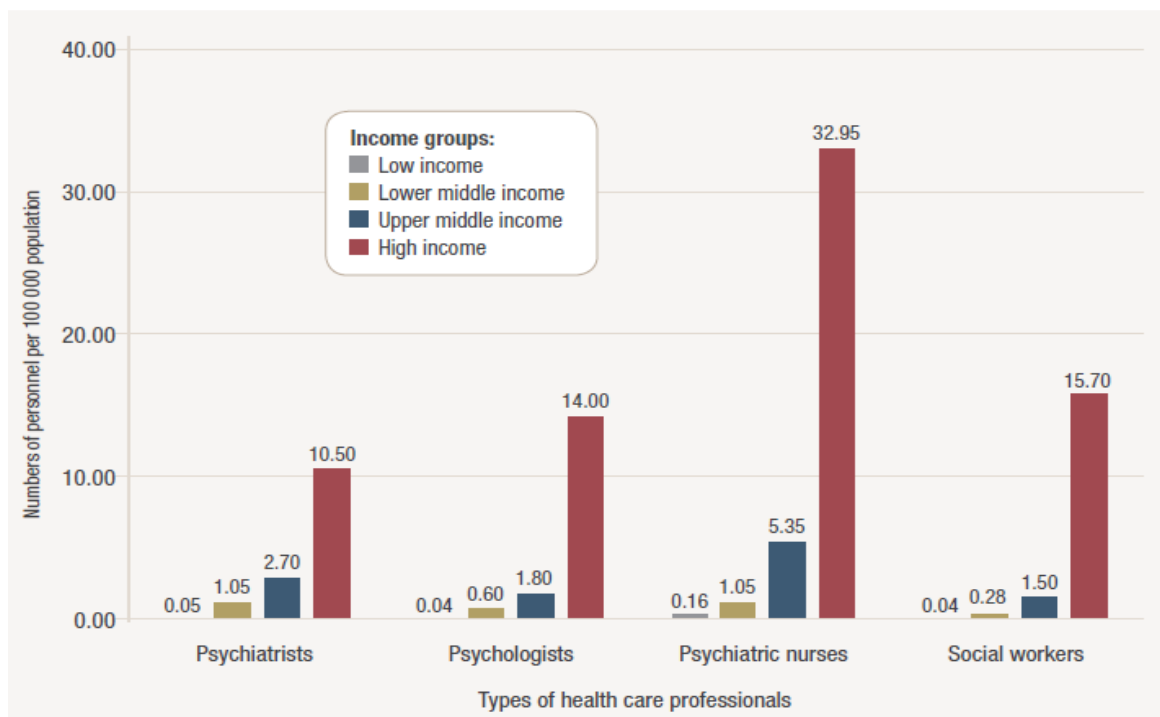


Figure 1.1: Disparities in mental health care services when comparing low- and middle income countries to high income countries [14]. *Reprinted with permission from the World Health Organisation. Accessed 26 August 2013.*

In conjunction with the issues associated with the lack of health care services for schizophrenia patients, it is important that the optimisation of antipsychotic treatment is considered. Although antipsychotic treatment is the most effective way to relieve the symptoms associated with schizophrenia and improve the daily functioning of individuals [15], this treatment remains effective in only a portion of individuals and is accompanied with many adverse drug reactions (ADRs) [16] (refer to Chapters 2.2.2 and 2.2.3 for more details). Thus, in order to further decrease the costs associated with schizophrenia by improving treatment outcomes, optimisation of treatment is required. It is for this reason that the implementation of pharmacogenetics, which is the study of how genetic variants influence treatments, is of utmost importance. Pharmacogenetics offers the opportunity to improve treatment outcomes, reduce ADRS and aid in the reduction of costs related to unnecessary hospitalisations and DALYs [17]. Considering that only 1% of drugs are developed for the treatment of diseases that affect individuals from poor countries [18], pharmacogenomics may also have a role to play in reducing the cost and time associated

with clinical trials in these countries. Using the example of the monoclonal antibody, trastuzumab, which is used in the treatment of cancer, the implementation of pharmacogenetic screening decreased the time required for the clinical trial from 10 years to 1.6 years and reduced the associated costs by an estimated \$35 million [19]. More directly, however, if genetic variants are identified to predict which patients will not respond to treatment or will develop ADRs, antipsychotic treatment regimes can be optimised prior to treatment. Some of the success stories with regards to the utilisation of pharmacogenetics for the optimisation of treatment include warfarin (*CYP2C9*, *VKORC1* and *CYP4F2*), clopidogrel (*CYP2C19*) and abacavir (*HLA-B*) [20]. The treatment outcomes of all three of these drugs have been shown to be influenced by specific genetic variants. The level of evidence for the role that these variants play is high and Clinical Pharmacogenetics Implementation Consortium (CPIC) dosing guidelines and Food and Drug Administration (FDA) pharmacogenomic drug label information are available for all three drugs [21,22]. With special reference to warfarin, the CPIC have designed a dosing algorithm which takes other factors such as age and weight into account along with genotype. The provided dose calculator allows the physician to enter values and provides warnings if the values entered are questionable. In the example provided in Table 1.1, the patient is a 70 year old Caucasian, 180 centimetres tall, weighs 74 kilograms and is not taking *CYP2C9* enzyme inducers or amiodarone. Furthermore, the patient is heterozygous for the *VKORC1* -1639 SNP and “wild type” for *CYP2C9*. Based on this information, a suggested dose of 29 mg/week was calculated [23].

Table 1.1: Warfarin dosing calculator. *Reproduced with permission from [23,24], Copyright Massachusetts Medical Society. Accessed 2 August 2013.*

Variable	Units or Allowed Values	Enter Value	Error Messages/Warnings
Age	Years	70	
Height	Centimetres	180	
Weight	Kilograms	74	
<i>VKORC1</i> genotype	A/A A/G G/G U (for Unknown)	A/G	
<i>CYP2C9</i> genotype	*1/*1 *1/*2 *1/*3 *2/*2 *2/*3 *3/*3 U (for Unknown)	*1/*1	
Race	A (for Asian) B (for Black or African American) C (for Caucasian or White) U (for Unknown or Mixed Race)	C	
Taking Enzyme Inducer	Y (for yes) or N (for no)	N	
Taking Amiodarone	Y (for yes) or N (for no)	N	
Computed Weekly Starting Dose (milligram/week):		29	

The utility of these findings is clear, however, there remain only a few examples where specific genotypes provide good evidence for predictive treatments. With the introduction of high throughput sequencing technologies and the ability to sequence entire genomes, there have been great advances in human genetics and genomics [25]. With these substantial developments, it is expected that the number of clinically useful pharmacogenetic findings will increase. This is highlighted by the fact that during the time when geneticists were feeling pessimistic about the lack of significant variants identified from genome-wide associations studies (GWAS), the field of pharmacogenomics was able to identify a number of significant findings, which were replicated [26,27]. Nonetheless the number of pharmacogenetic GWAS, with particular reference to whole genome and exome sequencing (WGES), remains limited [26]. Although association studies identifying novel pharmacogenetic findings through the use of WGES are currently lacking, the role for pharmacogenetics using this technology was recently highlighted in a study by Ashley *et al.* [28]. This study sequenced the genome of a healthy individual and by using the information available on PharmGKB (a web-based database which provides information about variants that have been associated with pharmacogenetic traits) was able to provide pharmacogenetic advice for 100 drugs. Furthermore, at the 2012 American Society of Human Genetics conference [29] the ethical considerations regarding the returning of results obtained from WGES studies, particularly those relating to incidental findings, was the topic of much concern. Nonetheless, it was almost unanimously agreed that the discovery of pharmacogenomic variants are of value to patients as they are likely to be clinically actionable [30]. Furthermore, the return of these results is not hampered by the stigma that may be associated with other incidental disease-related findings. In fact, the return of pharmacogenomic findings may aid in reducing stigma by improving the treatment of diseases and alleviating the associated symptoms. This is of particular relevance for schizophrenia, where the severe and debilitating symptoms are the cause of much stigmatisation and suffering [9]. Thus, the utility of WGES may be especially relevant to the field of antipsychotic pharmacogenomics. Not only are sequencing technologies likely to shed light on the unexplored areas of the genome, but they offer the opportunity to examine all variation simultaneously so as to provide a comprehensive picture of the variome. If these technologies can be utilised in LMIC where current research is limited and optimal treatments are urgently required, the high socio-economic burden of schizophrenia may be alleviated.

CHAPTER 2: Literature review

2.1 Schizophrenia

2.1.1 Clinical aspects of schizophrenia

Schizophrenia is a debilitating disorder affecting an estimated 24 million individuals across the globe [31]. This chronic and relapsing disorder, which is aggravated by an absence of cures and inefficient treatment [12], is considered one of the top ten causes of disease-related disability throughout the world [10]. In the USA, the severity of schizophrenia is reflected by the 100 000 beds that are occupied by schizophrenia patients every day, with the annual costs accounted to schizophrenia amounting to \$32.5 billion [32]. It is estimated that of all the individuals who develop schizophrenia, only 15% of patients will lead productive lives, while 60% will experience intermittent episodes and 25% will not be able to live independently [32]. This highlights the devastating nature of schizophrenia, which affects not only those living with the disorder, but places great emotional and financial strains on caregivers. Not only is the treatment of schizophrenia costly, but both affected individuals and their caregivers experience a substantial loss of productivity as a result of the disorder, with caregivers dedicating an estimated 6-9 hours per day to the care of schizophrenia patients [11].

The incidence and prevalence rates of schizophrenia are more or less consistent the world over, with about 8-40 new cases per 100 000 individuals reported annually, amounting to a lifetime risk of 0.7% [33]. Although the prevalence of schizophrenia is often reported to be approximately 1%, more recent estimates report a prevalence closer to 0.5% [34]. Nonetheless, this is a large portion of the global population that is affected by the disorder and although schizophrenia has been formally described for the last two centuries [34], the burden caused by the illness continues to affect individuals all around the world. The onset of schizophrenia usually occurs during adolescence or early adulthood [35], with age of onset ranging from 15-45 years of age, but most commonly occurring between 18-30 years, with an earlier age of onset correlating with more severe symptoms [36]. Substance abuse or a life stressor may precipitate the occurrence of schizophrenia, however, it is usually very difficult to pinpoint a specific event [36,37]. Schizophrenia is characterised by many severe and varying symptoms which can broadly be classified into positive, negative, cognitive and mood symptoms [36]. These symptoms are usually measured with the use of the Positive and Negative Syndrome Scale (PANSS) [38], the Scales for the Assessment of Negative and Positive Symptoms (SANS and SAPS) or the Brief Psychiatric Rating Scale (BPRS) [39]. More specifically the symptoms associated with schizophrenia are indicated in Box 2.1.

Box 2.1: The symptoms associated with schizophrenia [36]

- Positive Symptoms, which are absent in the general population, but present in schizophrenia patients and include:
 - Impaired reality testing
 - Delusions, which mostly occur as persecutory delusions and delusions of reference, but also include delusions of control, thought insertion, withdrawal and broadcasting
 - Hallucinations, of which auditory hallucinations are the most frequent
- Negative Symptoms, which are present in the general population, but absent in schizophrenia patients and include:
 - Abulia, which is a lack of motivation
 - Alogia, which is a poverty of speech
 - Anhedonia, which is the inability to experience pleasure
 - Avolition, which is a lack of initiative
 - Apathy, which is a lack of interest
 - Reduced social drive
- Symptoms of disorganisation, which include:
 - Formal thought disorder
 - Disorganised behaviour
- Mood symptoms, which often precede the formal onset of schizophrenia and include:
 - Impairment in affective experiences and expression
 - Increased emotional arousal
 - Depression
- Motor symptoms, which include:
 - Slowing of psychomotor activity
 - Excessive motor activity
- Cognitive symptoms, which include:
 - Impairment in episodic memory
 - Impairment in processing speed
 - Impairment in verbal fluency
 - Impairment in attention
 - Impairment in executive functions
 - Impairment in working memory

The progression of schizophrenia is lengthy and varies from patient to patient; however, it can be broadly divided into five main stages. These are the premorbid phase (stage 0), the prodromal phase (stage 1a and b), the first episode of psychosis (stage 2), incomplete remission/relapse (stage 3) and the persistent unremitting disease (stage 4) [40]. The features most commonly associated with the various stages are summarised below, however, it should be noted that it is not always easy to differentiate between these stages and the symptoms are often imprecise [36]. The premorbid stage is characterised by an increased risk for psychosis and subtle signs of cognitive difficulty and social isolation. The prodromal stage is associated with transient positive or general symptoms and a corresponding decline in function, with stage 1a exhibiting mild, non-specific symptoms and 1b exhibiting psychotic symptoms below the clinical threshold. Although individuals who are classified in this phase often go on to experience their first episode of clinical psychosis,

which marks stage 2, more than half of these “at risk” individuals do not develop schizophrenia. In order for a patient to be diagnosed as experiencing the first psychotic episode; hallucinations, delusions, disorganised speech or behaviour, or negative symptoms must be present for at least a month. At least one schneiderian (bizarre hallucination or delusion) or two core symptoms (hallucination, delusions, disorganised speech or behaviour, or negative symptoms) must be present [36,40]. However, the Diagnostic and Statistical Manual of Mental Disorders version five (DSM-5), released in 2013, has proposed that in all cases two core symptoms must be present before diagnosing a first episode of schizophrenia [41]. After experiencing the first episode of schizophrenia, the decline towards stages 3 and 4 is most pronounced in the first five years [36,40].

Schizophrenia is commonly referred to as a disorder, the most likely reason for this stemming from the lack of biochemical markers for diagnosis and the heterogeneity of the disorder, with symptoms varying from patient to patient [42]. Due to the occurrence of these varying symptoms, the DSM-IV [43] and International Classification of Diseases version 10 (ICD-10) [44] provide different sub-types of schizophrenia, however, it has been decided by the DSM-5 to remove these subtypes due to the overlap in many of the symptoms [41]. The previous sub-types that were included in the diagnosis of schizophrenia were (i) catatonic type, which refers predominantly to psychomotor symptoms; (ii) disorganised type, which refers to incoherent or inappropriate behaviours; (iii) paranoid type, which refers to hallucinations and delusions; (iv) schizoaffective type, which refers to mood and psychotic symptoms; (v) undifferentiated type, which refers to non-specific symptoms; and (vi) residual type, which refers to milder cognitive and negative symptoms. In addition to these six sub-types, the ICD-10 also refers to (i) simple type, which refers to the absence of severe delusions and hallucinations and (ii) latent type, which refers to the presence of odd behaviours [36,45]. Further complicating the understanding of schizophrenia is the fact that it can be very difficult to differentiate this highly heterogeneous disorder from other psychiatric disorders, as the symptoms overlap substantially [46,47] (Figure 2.1). If the heterogeneity of schizophrenia can be clarified and the boundaries between the psychiatric disorders can be made clearer, perhaps a more defined understanding of the disorder can be obtained to facilitate research efforts.

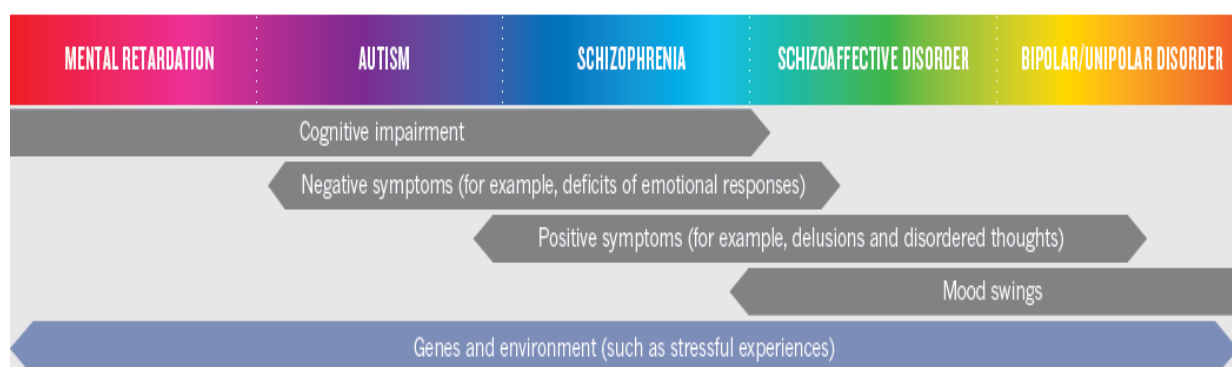


Figure 2.1: A representation of the overlap in symptoms observed between schizophrenia and other psychiatric disorders [47]. *Reprinted with permission from Nature Publishing Group.*

2.1.2 Risk factors for schizophrenia

Even though schizophrenia is complex and the lack of understanding of this disorder is heightened by the intricacy of the brain, there has been substantial research performed and several risk factors have been associated with the development of schizophrenia. These include (i) prenatal exposures such as stress, infection and poor nutrition; (ii) obstetric complications; (iii) childhood traumas; (iv) urban birth; (v) migrant status; (vi) adolescent cannabis abuse; (vii) brain injury; (viii) anti- N-methyl D-aspartate (NMDA) receptor encephalitis; (ix) male gender; (x) lower socio-economic class; (xi) winter/spring birth; (xii) famine; (xiii) poverty; and (xiv) older parental age [10,34,46,48]. Although these factors are associated with schizophrenia, the actual cause and effect remains difficult to elucidate in many cases. For example it is difficult to determine whether the abuse of cannabis causes the onset of schizophrenia or whether schizophrenia patients are more likely to abuse cannabis due to their disease status. Similarly, does migration and urban living precipitate schizophrenia, or do schizophrenia patients move towards urban areas? Does poverty act as a stressor for the development of schizophrenia or does the loss of employment as a result of schizophrenia lead to poverty? Furthermore, factors such as urban birth or migration may not directly cause schizophrenia, but may instead be linked to factors such as racism, poverty or a lack of social support networks [34,49].

Although the level of evidence for the above mentioned risk factors remains limited, one of the most robustly associated risk factors for the development of schizophrenia is genetics [48]. Individuals with two affected parents have a 45 fold higher risk of developing schizophrenia when compared to the general population [50]. Furthermore, when comparing monozygotic and dizygotic twins, the risk for developing schizophrenia is approximately four fold higher in monozygotic twins (40-48 vs. 10-17 fold higher risk of developing schizophrenia, when compared to the general population) [50]. Thus, a family history of schizophrenia, or shared genetic variation, remains the most reliable predictor for developing the disorder [48].

2.1.3 The genetics of schizophrenia

Schizophrenia has been reported to be the most heritable of the psychiatric disorders (65-81% heritability) [51,52] and for this reason there have been many studies performed attempting to elucidate the genes and variants involved in schizophrenia, with nearly 9 000 variants reported to be associated with the disorder in at least one study [53]. The studies identifying these genetic variants have developed substantially over the years, as technologies have advanced (Figure 2.2). Before the advent of GWAS in 2005, the majority of schizophrenia research was performed using linkage analyses or the examination of candidate genes. Linkage studies are reliant on utilising families affected with schizophrenia to identify regions of the genome that are shared between affected individuals, but not unaffected individuals, while candidate gene association analyses rely on some knowledge of the biological underpinnings of schizophrenia and investigate allele frequency differences in these genes between cases and controls [52,54,55]. The predominant biological pathways that are suspected to be involved in the development of schizophrenia include the dopamine, serotonin and glutamate pathways, due to the binding of antipsychotics to these receptors [56] and as such genes belonging to these pathways have been the focus of many candidate gene studies [53].

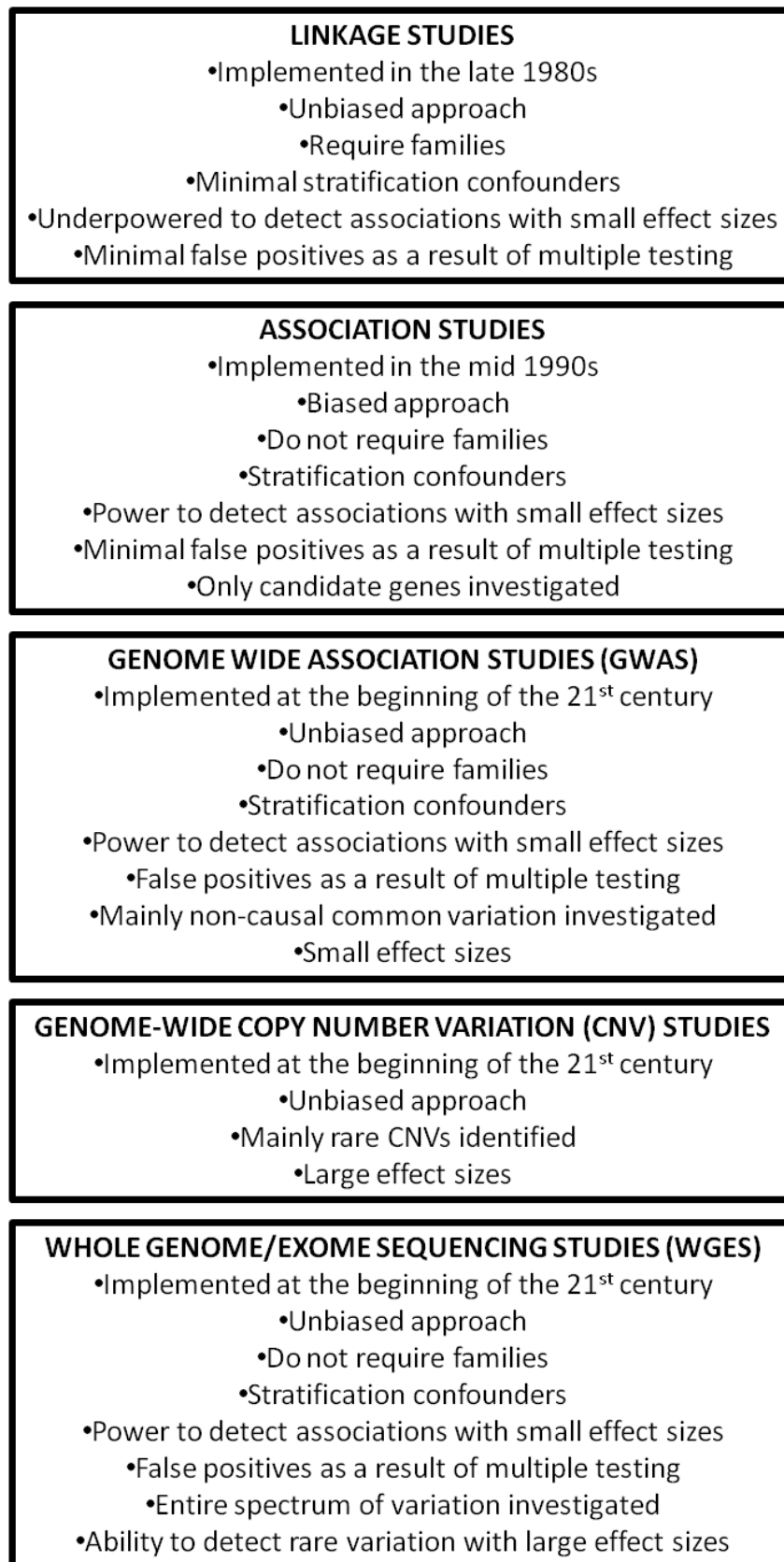


Figure 2.2: The progression of studies examining the genetic factors contributing to schizophrenia susceptibility [52,57].

Unfortunately, although there have been many promising results obtained from these candidate gene studies, the associations have been poorly replicated, with many conflicting results [42]. Furthermore, these studies have only provided information regarding biological aspects that we are already aware of. Thus, in conjunction with the GWAS era in human genetics, studies shifted their focus to large scale GWAS examining schizophrenia susceptibility [52,54,55]. GWAS investigating human disease were successful in identifying associations with other complex disorders such as type 2 diabetes, stroke, obesity and certain cancers [58] and thus the highly heritable and poorly understood psychiatric diseases were also expected to yield GWAS results that could help explain the missing heritability. At a similar time, studies also began to place their focus on scanning the genome for large structural variations or copy number variants (CNVs) [52,54,55]. Although the results obtained from these two approaches have not yielded conclusive results, there have been some interesting findings. The most noteworthy results that have been described with regards to the CNVs contributing to schizophrenia susceptibility include the 1q21.1, 2p16.3, 3q29, 15q13.3, 17q12 and 22q11.21 deletions and 1q21.1, 7q36.3 and 16p11.2 duplications, with the 22q11 deletion reported to be associated with a 25 fold increase in schizophrenia risk [42,52]. Interestingly, the results obtained from the GWAS and CNV studies have shown that common variants examined in GWAS have very small effect sizes, while rare CNVs seem to be highly penetrant and confer larger risks [34,59] (Figure 2.3).

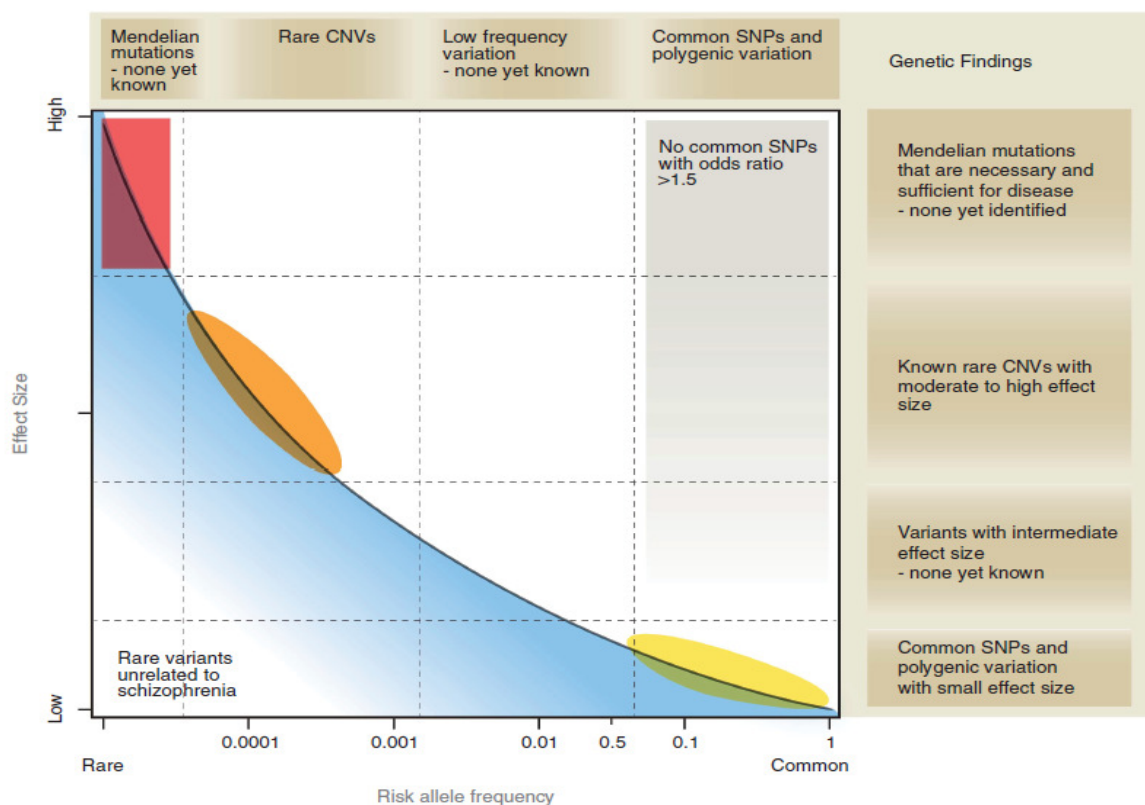


Figure 2.3: Relative contribution of common and rare variants to the development of schizophrenia. The yellow and orange ellipses represent associations that have been detected with common variants and rare copy number variants, respectively. The red rectangle and empty panel represent Mendelian mutations and intermediate frequency variants, for which no associations have been detected to date [59]. *Reprinted with permission from Nature Publishing Group.*

The advantage of GWAS lies in the fact that little is known regarding the biological mechanisms underlying schizophrenia; and as GWAS survey the entire genome, regions which are not known to be involved in schizophrenia may be uncovered. To date, the most replicable result obtained from the published schizophrenia GWAS, relates to variants occurring within the major histocompatibility complex (MHC). This region of the genome is involved in immune function and variants present in this area corroborate the theory that exposure to infection during gestation may increase the chances of developing schizophrenia [48,52]. Although, these results do make sense in the context of the biology of schizophrenia, the high gene density and low linkage disequilibrium (LD) of this region make identifying causal variants very difficult [52]. Unfortunately, to date, other findings from schizophrenia GWAS have not generated many replicable results [48]. The lack of significant and reproducible results obtained from these GWAS may in part be attributed to insufficient sample sizes. The millions of SNPs that are genotyped in GWAS result in multiple testing issues and in order to obtain P values that meet the requirements for genome-wide significance ($P \leq 5 \times 10^{-8}$) thousands of samples are required, especially in cases such as psychiatric disorders where the effect sizes are expected to be small. For this reason schizophrenia consortiums are required in order to pool patient cohorts and increase sample sizes. The largest of these consortiums is the psychiatric genome consortium (PGC), which is a multi-national effort to combine a large number of schizophrenia patients for GWAS in order to allow for the identification of more robust associations [60]. In the latest PGC schizophrenia GWAS publication a three stage strategy was utilised, which consisted of (i) a discovery cohort of 5 001 schizophrenia cases and 6 243 controls, (ii) a meta-analysis stage utilising 8 832 cases and 12 067 controls and (iii) a replication cohort of 7 413 cases and 19 762 controls, making this the largest schizophrenia GWAS to date [61]. This study identified 22 significantly associated loci, of which 13 were novel, and implicated the neuronal calcium signalling pathway in schizophrenia susceptibility. In addition, some of the findings from the previous PGC GWAS were highlighted once again. These included significantly associated variants in the vicinity of (i) one of the most well replicated findings in schizophrenia GWAS, namely the MHC complex, (ii) *MIR137* (a microRNA) and (iii) *TCF4* (a neuronal transcription factor). These findings emphasise once again not only the role that the MHC may play in schizophrenia susceptibility, but also the importance of epigenetic and transcriptional regulation, as well as gene networks [54,61,62].

Unfortunately, even with the impressive sample sizes that were utilised by the PGC, the effect sizes of the variants that were associated in these studies remain small, with odds ratios of only between 0.8 and 1.2 and associated variants only explaining a portion of the heritability [61]. The results from the published GWAS highlight the likelihood that hundreds of variants and genes may be involved, each conferring a small risk for the development of schizophrenia [63]. However, as GWAS genotype common variants and rely on brute force to identify associations, it is possible that the techniques utilised are unable to detect the effects of rare causal variants. Furthermore, as the associated variants are mostly intergenic, it is difficult to interpret the biological significance of such findings [55]. Therefore, with the elegant publication by Ng *et al.* [64] describing the ease with which the gene associated with Freeman–Sheldon syndrome was identified, there was much anticipation for the application of high throughput sequencing for other diseases. Even though the study by Ng *et al.* [64] examined a rare Mendelian disorder, the likelihood that WGES in the context of schizophrenia may uncover rare causal variants with large effect sizes, was met with

anticipation. For this reason schizophrenia studies have recently begun to shift their focus towards the use of high throughput or next generation sequencing (NGS). The use of exome sequencing has allowed the focus to shift from variants that have no known affect to variants that occur within coding regions, thus increasing the likelihood that causal variants may be directly identified.

Although it appears likely that NGS technologies will be the focus of future research, to date there have only been six published studies using these technologies to investigate schizophrenia susceptibility, of which two examined the South African Afrikaner population [65,66]. Four of these six studies focussed on sporadic cases of schizophrenia and identified *de novo* variants present in probands by sequencing trios or quads with only one affected offspring [65–69]. By utilising this approach they could overcome the issues associated with the overwhelming amount of data that is generated by WGES [65,66,68]. All four of these studies found that there was a higher rate of *de novo* variants in affected individuals and that these *de novo* variants were more likely to be non-synonymous and/or adversely affect the protein product [65,66,68,69]. Furthermore, the study by Gulsuner *et al.* [69] highlighted the role that damaging *de novo* variants played in disrupting genes involved in the neurogenesis of the prefrontal cortex. The fifth study also utilised a family based approach and sequenced representatives from five large families that were affected by schizophrenia. This study identified protein altering variants in genes associated with the NMDA receptor gene, which may be implicated in schizophrenia [70]. The last WGES study focussed on the identification of rare variation, which may have been overlooked by GWAS. This study utilised a discovery cohort of 166 schizophrenia patients (10% African, 89% European and 1% other), which were enriched for treatment resistance and a family history of schizophrenia, and 307 controls (6% African, 92% European and 2% other). The genomes/exomes of these individuals were sequenced and rare variants, which were predicted to alter the protein products, were identified. A total of 5 155 variants which were present in one or more cases and no controls were identified. These variants were subsequently prioritised for genotyping in 2 617 cases (45% African, 54% European and 1% other) and 1 800 controls (37% African, 59% European and 3% other). Association analyses did not reveal any variants that reached genome-wide significance, however, 49% of the variants were detected in only one or two schizophrenia individuals highlighting the possibility that rare functional variation may be involved in the development of schizophrenia [67]. Furthermore, as the majority of variants which occur at an allele frequency of less than 0.5% are only present in one population group [71], these findings may emphasise the importance of sequencing the genomes of schizophrenia patients and controls from many different populations.

Although genetic studies examining schizophrenia susceptibility continue to develop, there remains limited knowledge regarding the exact role that genetics plays. This may in part be attributed to locus heterogeneity, incomplete penetrance and the influence of the environment [42]. It remains likely that the variants contributing to schizophrenia differ between unrelated individuals, and that although there may be a common causal pathway, due to the heterogeneity of the variants and genes involved, GWAS may be unable to detect these pathways. Furthermore, due to the genetic persistence of schizophrenia, it has been postulated that the variants involved in the precipitation of schizophrenia are maintained throughout generations due to some evolutionary advantage and that the schizophrenia

phenotype is only expressed above a certain threshold, which is no longer advantageous [72]. This highlights the likelihood that many variants may be involved in schizophrenia. Lastly, due to the current inability to explain the development of schizophrenia by examining genetic or environmental factors as separate entities, it seems likely that the two may interact [34], suggesting a role for epigenetics in schizophrenia susceptibility (refer to Chapter 7.3 for further details). The current evidence, thus, points to the probability that schizophrenia is a complex multi-factorial and polygenic disorder [73]. Therefore, in order for genetic studies to be successful, WGES to examine all variation simultaneously in combination with well characterised cohorts may be required.

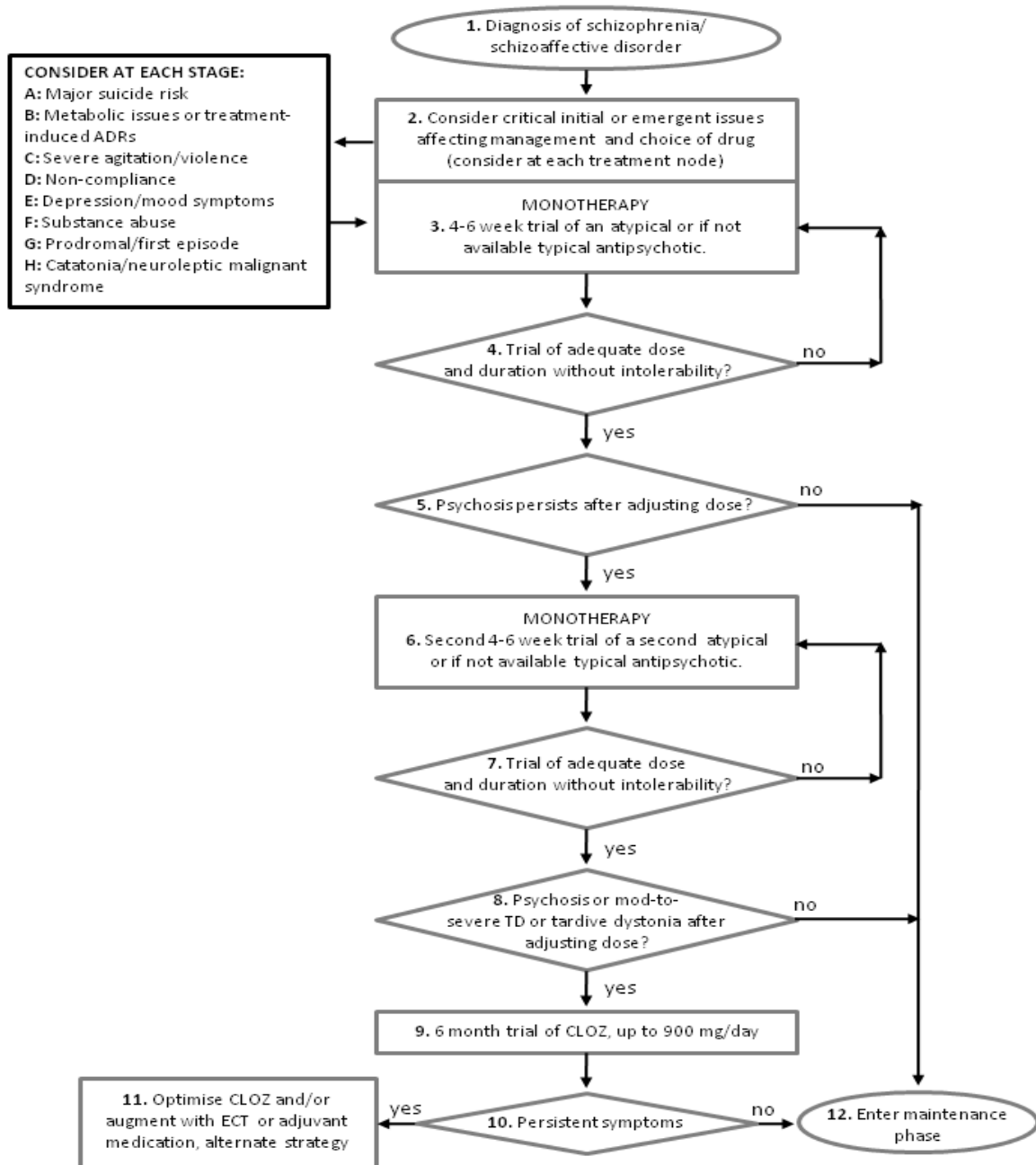
2.2 The antipsychotic treatment of schizophrenia

2.2.1 Background

Central to the management of schizophrenia is the successful treatment of the disorder. Until the 1950s, with the opportune discovery of the first antipsychotic, chlorpromazine [74], institutionalisation of patients was the treatment option of choice and only 26% of patients were discharged [75]. The doubling in the improvement of patients with schizophrenia in the 1970s, when compared to the 1920s, is largely attributed to the introduction of antipsychotics in the 1950s [76]. The introduction of these agents into the treatment of schizophrenia is arguably the greatest advancement in schizophrenia research. To date, antipsychotics remain the only effective therapeutic agents that are available for the treatment of schizophrenia [15] and their effects have consistently been shown to be superior to placebos [77–79]. Highlighting the value of these drugs, the WHO currently lists four antipsychotics (chlorpromazine, risperidone, fluphenazine and haloperidol) as essential medications [80], which are thus described as “of utmost importance” and “indispensable and necessary for the health and needs of the population” [81]. The importance of antipsychotic research is reflected by the 3 782 articles relating to the search term “antipsychotic” that were recorded in the year 2012 in the PubMed Database (accessed 2 August 2013) [82].

Antipsychotics mainly treat positive and disorganised symptoms and prevent relapse, however, they are less effective for cognitive or negative symptoms [15,83]. In addition to the inability of antipsychotics to treat all symptoms, it has been reported that approximately 40% of patients do not respond substantially to medication and up to 70% develop ADRs [84,85]. Furthermore, these differences in treatment response are varied and include differences in the time, dose and medication required for response, as well as differences in the development of ADRs. Thus, patients need to be considered individually and each case requires careful monitoring, making the treatment of schizophrenia complicated. Treatment algorithms are often incorporated in order to aid in the treatment of a patient based on the response that is observed. Unfortunately, obtaining the desired response may require the use of multiple different antipsychotics and this trial-and-error based process may take months (Figure 2.4) [86,87]. When making decisions on how to treat a patient, it is important to consider factors relating to the patient (e.g. age, gender and history of treatment response), the illness (e.g. duration, symptom type and comorbidity), the medication (e.g. efficacy, tolerability and cost) and the environment (e.g. diet, support and financial situation) [83]. Furthermore, if non-optimal treatment outcomes are observed,

decisions need to be made regarding whether to change the type or the dose of medication. According to consensus recommendations, patients showing little or no response to first line treatment should be treated for 3-6 weeks before changing the treatment, and patients showing partial response should be treated for 4-6 weeks. Thereafter, the dose should be increased or an alternate antipsychotic should be used [88]. Generally, if the patient shows an inadequate response coupled with ADRs, the medication is changed, whereas when an inadequate response without ADRs is observed, the dose is changed [88]. Ideally patients should respond quickly to antipsychotic medication with no ADRs. Unfortunately this is seldom the case [89].



© Copyright 2004-2006 International Psychopharmacology Algorithm Project (IPAP).

ADRs: Adverse drug reactions, TD: tardive dyskinesia, CLOZ: clozapine, ECT: electroconvulsive therapy.

Figure 2.4: International Psychopharmacology Algorithm Project Dosing Algorithm [86].

As mentioned previously, the first antipsychotic to be discovered was chlorpromazine (in the 1950s) and the introduction of this treatment allowed for the deinstitutionalisation of patients in the 1960s, which was a leap forward in the treatment of schizophrenia patients. Since then an additional 50 first generation antipsychotics (FGAs) have been introduced, however, treatment with many of these drugs is predominantly accompanied with extrapyramidal symptoms (EPS) [15]. In the 1960s, there was another prominent advance in the treatment of schizophrenia with the development of clozapine, the first second generation antipsychotic (SGA) [90]. Clozapine showed a greater efficacy when compared to the FGAs and was not accompanied with EPS. However, treatment with clozapine was in some instances accompanied by a severe and potentially lethal ADR known as agranulocytosis [90]. Due to the presence of this serious ADR, clozapine was not widely used until the 1990s when the superiority of clozapine with regards to response in treatment refractory patients and the reduction of suicide was demonstrated [91,92]. Even though clozapine remains superior to other antipsychotics in terms of efficacy, it is still not used in the first line treatment of schizophrenia in most settings [93]. Coinciding with the more accepted use of clozapine in the 1990s, a second round of SGA development took place, with the hope of introducing antipsychotics that would be as effective in reducing symptoms as clozapine, but unaccompanied by agranulocytosis. To date, there are twelve available SGAs, excluding clozapine, however there appear to be no clear differences in the treatment response profiles of FGAs and SGAs [15]. Although SGAs are associated with less EPS, they are unfortunately accompanied with a greater rate of weight gain and related ADRs [94]. To corroborate these findings, the Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia (CUtLASS), Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE), Schizophrenia Outpatients Health Outcome study (SOHO), Comparisons of Atypicals in First Episode (CAFE) and European First Episode Schizophrenia Trial (EUFEST) studies have examined various aspects of antipsychotic treatment [95]. Although these studies varied in the approaches used, the general message obtained was that antipsychotics are heterogeneous in nature, the efficacy of different antipsychotics is similar, with the exception of clozapine which appears to be superior, and FGAs are more frequently associated with EPS, while SGAs are more frequently associated with weight gain and resulting metabolic conditions [95].

Although many gaps remain in our understanding of the mechanism of action of antipsychotics, there are certain aspects that have been elucidated. The central mechanism of action for all classes of antipsychotics is their binding to dopamine D₂ receptors [96]. In addition, SGAs bind not only to dopaminergic receptors, but also to serotonergic receptors [97]. Although the dopamine and serotonin receptors are the most well documented targets for antipsychotics, this class of drugs may also bind to alpha-adrenergic, histaminic and muscarinic receptors [96,98]. The different affinities that antipsychotics have for receptors and the differences in their absorption, distribution and elimination profiles contribute to the varying treatment response profiles that are observed between individuals and drug types [15,96]. Nonetheless, the core feature of antipsychotics remains their ability to reduce the availability of dopamine, which results in a decrease in positive symptoms (psychosis, mania, tics, aggression). Unfortunately, as mentioned previously, this mechanism of action does little to reduce negative or cognitive symptoms and the reduction in dopamine can in fact result in the aggravation of these symptoms [96]. Thus, the goal of optimal treatment with current antipsychotics would be to improve their regional specificity by allowing for an

antagonistic effect on the mesolimbic dopamine, to alleviate the positive symptoms, and an agonistic effect on frontostriatal dopamine to prevent the occurrence of negative and cognitive symptoms [83]. Therefore, research into the development of new antipsychotics is driven by the need to treat negative and cognitive symptoms in addition to the positive symptoms. Research initiatives such as the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) study focus on targeting glutamatergic, cholinergic, gaba-ergic, neuroleptic, cannabinoid and non-neurotransmitter receptors, in addition to the dopamine and serotonin receptors [56]. However, to date results obtained from these initiatives are limited.

In addition to antipsychotic therapy, other options for treatment exist. These include (i) electroconvulsive therapy, which is utilised mainly for the treatment of catatonic symptoms; (ii) repetitive transcranial magnetic stimulation, which is employed chiefly for the treatment of positive symptoms and auditory hallucinations; (iii) deep brain stimulation; (iv) family and patient psychoeducation, which aids in improving the affected individuals and their caregivers' understanding of the disorder, while also providing coping strategies; (v) cognitive behaviour therapy that provides patients with the ability to rationally explain their symptoms; (vi) social skills training, which aids in improving the daily living skills of the patients; and (vii) assertive community treatment, which utilises a multi-disciplinary approach to deliver clinical care to patients [15]. Ideally, all patients should receive, in addition to antipsychotic medication, psychosocial interventions. As may be noted, many of these treatment options are focussed on the symptoms that are not treated by antipsychotics and are used in combination with the therapeutic agents. They also serve as reminder of the deficits with regards to antipsychotic treatment, which currently requires optimisation [73].

2.2.2 Response to antipsychotic treatment

It is difficult to define treatment response, which like schizophrenia, is a highly heterogeneous phenotype, with individuals responding differently to antipsychotic medications. This difference in response is reflected in the British National Formulary which states that "various antipsychotic drugs differ somewhat in predominant actions and side effects. However, the differences between antipsychotic drugs are less important than the greater variability in patient response" [16]. If the biological mechanisms underlying the differences in treatment response phenotypes can be determined, treatment outcomes can be optimised in such a way that individuals can be treated according to their response profiles. Although it is difficult to compare the differences in response observed between different patients, there are a few main aspects that should be considered when examining treatment response outcomes. The factors that are central in response to antipsychotic treatment include treatment resistance, remission and relapse (Figure 2.5).

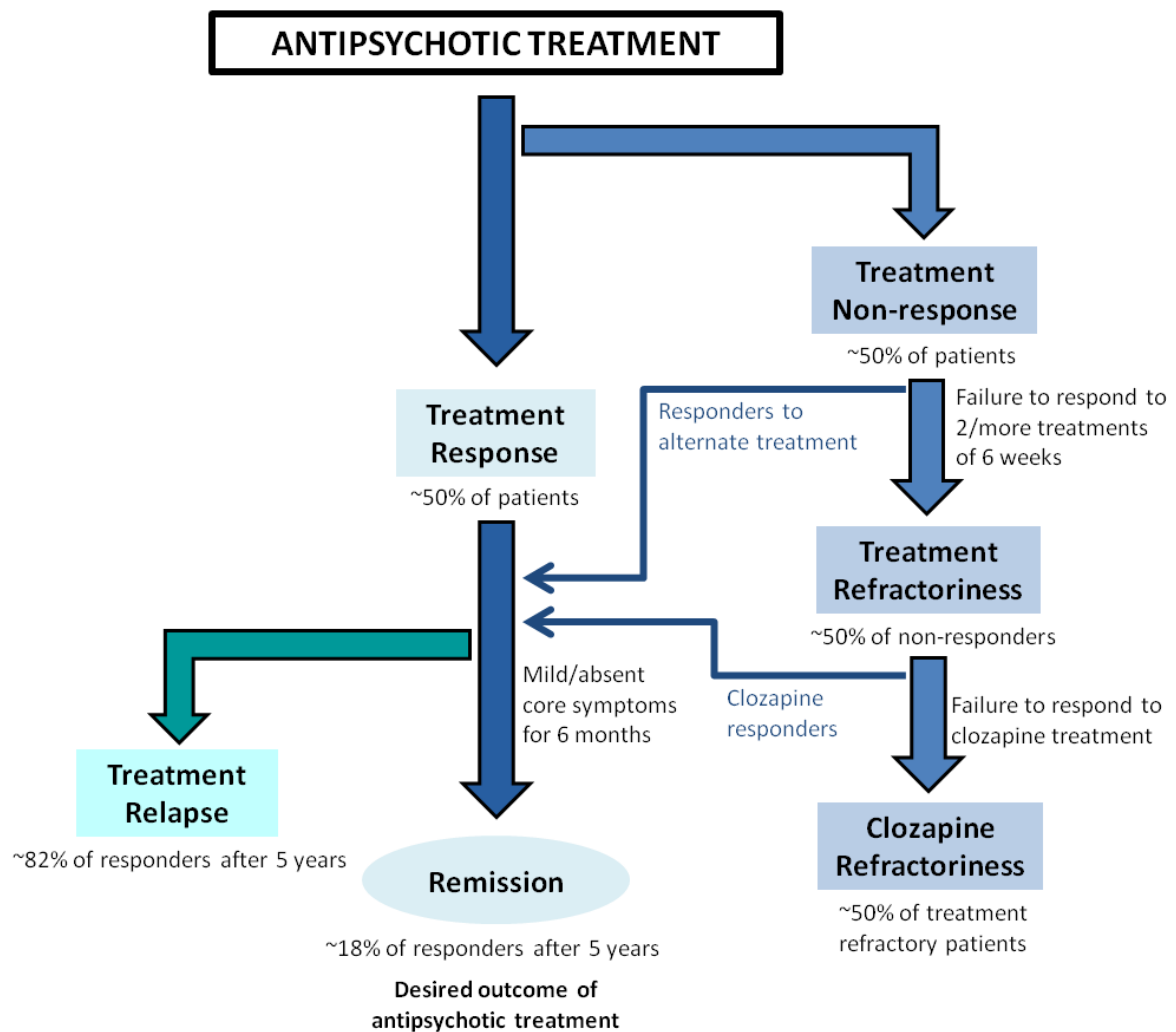


Figure 2.5: The central features of response to antipsychotic treatment [84,99–101].

A patient is usually considered resistant to treatment or treatment refractory after 2-3 successive treatments have been prescribed without the alleviation of symptoms, with particular reference to positive symptoms [102]. In more robust definitions, the use of clozapine is typically considered an indication of treatment resistance [67]. As mentioned previously antipsychotic treatment is only effective in a subset of individuals and treatment resistance occurs in about 30-50% of patients [46,103]. Although the study of treatment response is vital for the effective treatment of schizophrenia, the phenotype is difficult to define and there have been several different methods of identifying treatment refractory patients and defining non-response. In earlier studies a less than 20% improvement in PANSS/BPRS scores was used to identify patients that were not responding to treatment. However, more recently, guidelines for the identification of treatment refractory patients have suggested that an individual be considered treatment resistant if pervasive symptoms remain present after treatment with 1-3 different antipsychotics, with each treatment lasting 4-6 weeks. The presence of pervasive symptoms is measured in terms of a score of greater than 4 on the Clinical Global Impression Scale, which ranges from 0-7 (equivalent to 75 on the PANSS scale or 45 on the BPRS scale); a score greater than 49 on the Functional Assessment for Comprehensive Treatment of Schizophrenia scale and a score greater than 50 on the Global Assessment of Functioning scale, both of which range from 0-100 [102]. In

contrast to treatment resistance, remission without relapse is the ultimate goal of the antipsychotic treatment of schizophrenia. As with other aspects of treatment outcomes, remission is difficult to define, however, the Remission in Schizophrenia Working Group defines remission as the absence, or mild presence, of eight core schizophrenia symptoms (delusion, unusual thought content, hallucinatory behaviour, conceptual disorganisation, mannerisms/posturing, blunted affect, passive/apathetic social withdrawal and lack of spontaneity and flow of conversation) for at least six months, while partial remission does not require the six month time period to be included in the definition. These symptoms should be assessed by means of the PANSS or SANS/SAPS scales [104].

As mentioned above it is the goal of antipsychotic treatment to allow patients to reach remission and many patients do reach this goal. However, unfortunately the majority of patients will relapse and experience additional psychotic episodes or a worsening in the observed symptoms [105]. It has been reported that the majority of first episode schizophrenia (FES) patients respond well to treatment in the first year, with 80% of patients obtaining remission status and only 16% of patients relapsing. Unfortunately, by year five 82% of the patients relapse [101]. Examining patients after fifteen years, it has been reported that only 0-12% of patients remain relapse free [106]. Thus, although patients who are treated early do for the most part improve after short-term treatment, long term treatment is associated with relapses, ongoing symptoms and a poor quality of life for patients [56,107]. Furthermore, the majority of costs associated with schizophrenia can be attributed to the relapse of patients, with an estimated \$2 billion dollars spent per annum in the USA on the readmission of schizophrenia patients [108]. Although, to our knowledge, the costs related to relapse in South Africa are not known, due to the lack of resources allocated to the treatment of psychiatric disorders in LMIC (refer to Chapter 1) [3,4], it seems likely that these costs will also be high. It has also been estimated that the mental health costs of relapsed patients are 2-5 times higher than non-relapsed patients [109]. Therefore, by preventing relapse, the costs associated with schizophrenia can be greatly reduced.

In addition to examining the occurrence of relapse, Robinson *et al.* [101] reported that patients who discontinue medication were at a five-fold higher risk for relapse. Unfortunately, the measurement of non-adherence has proven to be problematic, as self-reporting is very unreliable. Furthermore, although treatment outcomes and response rates are better in patients experiencing their first episode of psychosis [110,111], non-adherence is also highest in this group, with 39% of FES patients discontinuing medication and only 20% remaining adequately adherent [110]. Thus strategies to improve adherence rates are urgently required. In a recent survey, 84% of psychiatrists referred to lack of insight, belief that patients no longer require medication and ADRs as the most likely reasons for non-adherence, highlighting the need for optimised treatment protocols with reduced ADRs in order to prevent discontinuation [112]. The case study in Box 2.2 provides real life evidence for the effect of ADRs on adherence and treatment outcome and the corresponding need for optimised treatments [113].

Box 2.2: Case study 1

A highly functioning individual experienced her first episode of psychosis after which she was prescribed a low dose of antipsychotic medication. Ten weeks after initiation of the antipsychotic treatment, her symptoms disappeared. However, the disappearance of the symptoms was accompanied with sedation and extrapyramidal side effects, which she did not want her peers to find out about. She subsequently stopped treatment and the hallucinations and delusions returned. These symptoms were more intense than those experienced in her first episode of psychosis and re-initiation of treatment was less effective than before [113].

One of the most effective strategies to combat non-adherence is the use of long-acting injectables (LAI). When comparing LAI to oral medication it has been reported that in the LAI group there were fewer discontinuations (26% vs. 70%), more symptoms reductions, higher remission rates (64% vs. 4%) and lower relapse rates (9.3% vs. 42%) [114]. Although LAI were introduced in the 1960s and have been shown to result in better treatment outcomes, less than 20% of patients receive LAI and less than 10% of psychiatrists offer LAI as a first line treatment [115]. Despite the fact that LAI are not commonly prescribed, 62% of psychiatrists reported that the use of a LAI would be the best way to reduce non-adherence [112]. As early treatment is associated with better outcomes and FES patients are most likely to be non-adherent, the use of LAI during this time frame may be optimal. By providing early intervention and improving adherence in these patients, it seems likely that better treatment outcomes can be attained. LAI are usually not prescribed due to their association with a lack of autonomy for the patient. However, if the psychosis is not alleviated the patient will not be able to make independent decisions. An example of how patients may benefit and even prefer the use of LAI is demonstrated by the case study reported in Box 2.3 [116].

Box 2.3: Case study 2

A woman experienced persecutory delusions and was thus hospitalised due to her destructive and aggressive behaviour. She was subsequently prescribed 5 mg/day of oral risperidone after which she experienced abnormal involuntary movements and was therefore prescribed olanzapine. One day after she was discharged, she decided to discontinue her medication and four days later she was re-hospitalised due to physical aggression. She was then prescribed 20 mg/week of flupenthixol LAI. The voices that she was hearing finally disappeared and she no longer felt as if she was being persecuted. Twelve months after remission she experienced some movement disorders and was prescribed risperidone LAI. She personally opted to remain on the LAI, rather than oral treatment, as she found it more convenient and was worried about relapsing again [116].

In addition to non-compliance, other factors have also been associated with poor treatment outcomes. Poor treatment outcomes have been associated with (i) male gender; (ii) early age of onset; (iii) early response outcomes, with response over 2-4 weeks highly indicative of the long term response; (iv) ethnicity; (v) co-morbidity; (vi) prolonged periods of untreated illness; (vii) severity of cognitive and negative symptoms (viii) lack of insight; (ix) poor pre-morbid functioning; and (x) a family history of schizophrenia [10,15,103,117,118]. With special reference to duration of untreated psychosis (DUP) as a predictor for treatment

outcomes, it has been reported that 87% of patients experiencing their first episode of psychosis will respond to treatment in comparison to 20-30% of multi-episode patients responding to treatment [107]. It should however be noted that those patients responding well to treatment may be filtered out of multi-episode studies due to the fact that they no longer require hospitalisation or treatment [113]. Thus, in order to obtain maximum treatment efficacy, many factors need to be taken into account. Most importantly early and optimal treatment protocols appear to result in the better treatment outcomes.

2.2.3 Adverse drug reactions caused by antipsychotic treatment

When examining antipsychotics as a whole, they are associated with many debilitating and varied ADRs. These include neurological, cardiovascular, gastrointestinal, haematological, genitor-urinary, musculoskeletal and endocrine ADRs [15]. These varied and severe ADRs are a major reason for non-compliance [119]. Thus, if the mechanisms underlying the occurrence of ADRs can be determined, compliance and subsequently treatment outcomes can be improved. Although ADRs are not completely understood, in general terms, the occurrence of ADRs is related to the interference of antipsychotics with different pathways. Although the mechanism of action of antipsychotics relies on the blockade of certain receptors, excessive blockade of these or other receptors may result in ADRs [98]. The unwanted effects related to the antagonism of various receptors caused by antipsychotics include excessive blockade of: (i) dopamine receptors in the nigrostriatal pathway, which may result in EPS; (ii) dopamine receptors in the hypothalamic pathway, which may result in hyperprolactaenemia; (iii) dopamine receptors in the mesocortical pathway, which may aggravate negative and cognitive symptoms; (iv) histaminergic receptors, which may result in sedation, weight gain and metabolic disorders; (v) cholinergic receptors, which may result in a dry mouth, constipation, impaired cognition, blurred vision, urinary retention and tachycardia; and (vi) alpha1-adreniergic receptors, which may result in orthostasis [83,120].

The differences in ADRs that are observed between antipsychotics are related to the differences in binding profiles that these drugs display. When examining the 64 antipsychotics that are available on the market, their ADR profiles vary widely [15]. However, in broad terms, the ADRs related to FGAs can be grouped separately to those occurring as a result of SGAs [83,96]. FGAs have a greater affinity for dopamine receptors and are thus more likely to result in movement disorders and changes in prolactin levels, with tardive dyskinesia (TD) occurring in approximately 20-30% of FGA treated patients [121] and antipsychotic sexual dysfunction as a result of changes in prolactin levels occurring in 30-80% of patients [119]. The occurrence of these ADRs can be attributed to the narrow therapeutic range of antipsychotics, with 60-80% striatal dopamine receptor occupancy required in order for the antipsychotics to reduce symptoms. However, if a greater than 80% dopamine occupancy occurs, the above mentioned movement and prolactin related ADRs are more likely to occur [97,122]. One of the most severe movement disorders is TD, which typically develops about three months or later after treatment. TD is a movement disorder which is characterised by repetitive and involuntary movements which occur predominantly in the orofacial region, however, may also include the limb and trunk regions. The observed movements include chewing, tongue protrusion, lip smacking, puckering and pursing of lips and rapid eye blinking [123]. TD is usually diagnosed using the Schooler-Kane TD Research and Diagnostic Criteria [124]. EPS ADRs are less severe than TD

and occur within the first days or weeks of treatment. This group of ADRs include dystonia (sustained abnormal posture), akathisia (inner unrest) and parkinsonism (tremors, rigidity and bradykinesia) [125]. With regards to the changes in prolactin levels, an increase of 2-10 fold in prolactin has been observed in patients receiving FGAs [86]. This increase in prolactin levels has been reported to result in an increase in breast cancer as well as gynaecomastia, galactorrhoea, amenorrhoea, sexual dysfunction and osteoporosis [56,126].

Due to the fact than SGAs have a high affinity for serotonin receptors, the dopamine receptors are less likely to be blocked, and therefore less EPS are observed in relation to these drugs [83,96]. However SGAs, unlike the FGAs, have a higher affinity for the serotonergic and histaminergic receptors [83] and thus a higher rate of weight gain is observed, with about 30% of SGA treated patients gaining weight [121]. Although weight gain in itself is a serious ADR, it is often accompanied with various other factors which can diminish the quality of life of these patients. These factors include diabetes, hyperglycemia, hyperlipidemia, hyperphagia, cardiovascular disease, depression, low self-esteem and non-adherence to medication [121,127]. When comparing schizophrenia patients to controls, it was observed that schizophrenia patients had three times as much intra-abdominal fat [128]. Furthermore, after treatment with SGAs, it has been reported that the prevalence of obesity increases by 10% [129]. This is of serious concern due to the rise of obesity in the general population, with 68% of the general US population considered overweight and 34% considered obese [130]. As the rate of obesity and type II diabetes is even higher in schizophrenia patients, this may have severe health consequences for schizophrenia patients [131–133]. Highlighting these health concerns, it has been reported that 50% of schizophrenia patients develop metabolic abnormalities, which are accompanied with increased cardiovascular risk and mortality [131]. Antipsychotic induced weight gain, although poorly understood, can be attributed to imbalances in energy intake (e.g. appetite and food intake) and energy expenditure (e.g. decreased activity or metabolism) [134].

Another drug which differs notably with regards to its ADR profile is clozapine, which causes agranulocytosis in 0.8% of cases [135]. Agranulocytosis occurs as a result of the destruction of defensive white blood cells and this may result in fatality in approximately 10% of cases [136]. As such, due to the agranulocytosis related deaths of eight clozapine-treated Finnish schizophrenia patients in 1975, the use of clozapine was curtailed [137]. However, in the 1980s a study showed the superiority of treatment utilising clozapine in treatment refractory patients [91]. This was an important finding due to the fact that as many as 25% of schizophrenia patients are treatment refractory and 50% of these patients respond to clozapine [99,100]. Therefore in the 1990s clozapine was once again introduced globally. In order to prevent the occurrence of agranulocytosis and to reduce fatality rates, patients now undergo routine hematologic monitoring. This monitoring includes weekly white blood cell counts for the first 18 weeks, after which the white blood cells are counted every four weeks [121,138].

As has been highlighted thus far, antipsychotics, in combination with other treatment strategies, can prevent relapse and reduce symptoms and in so doing reduce the disease burden associated with schizophrenia. However, antipsychotics may be costly and can result in ADRs and poor treatment efficacy, which increase the burden of disease (as described above and summarised in Figure 2.6). It is thus important to ensure that the benefits

associated with antipsychotics outweigh the negatives. If treatment can be optimised to improve treatment response outcomes and eliminate ADRs, the overall burdens associated with the disorder can be decreased.

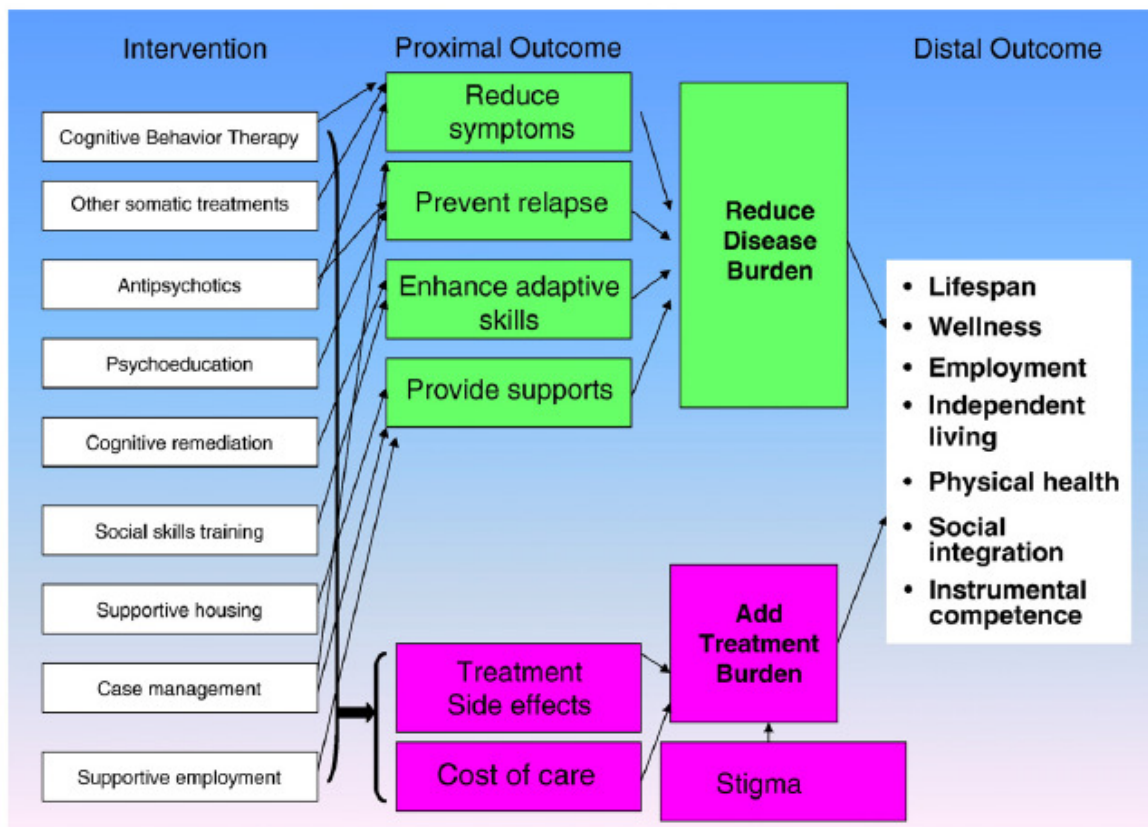


Figure 2.6: The role that treatment interventions have on reducing the burden of schizophrenia [15]. *Reprinted with permission from Elsevier.*

2.3 Antipsychotic pharmacogenomics

2.3.1 Background

At the same time that the first antipsychotic drug was introduced in the 1950s [74], the term pharmacogenetics was coined to refer to the influence that genetic variants may have on the differences in drug responses that are observed between individuals [139]. Interestingly, in the 1990s when the second generation of antipsychotic drugs were introduced [15], technologies to scan the entire genome were developed, and with these new approaches, the term pharmacogenomics came into use. Pharmacogenomics refers to the analysis of variants in all the genes in the genome rather than focussed candidate approaches. However, as genome-wide analyses are becoming common practice, the terms pharmacogenetics and genomics are used interchangeably [140,141]. Pharmacogenomics can further be sub-divided into efficacy pharmacogenomics, which examines genetic variants contributing to differences in treatment response, and safety pharmacogenomics, which examines how genetic variants contribute to the development of ADRs [140].

In terms of antipsychotic pharmacogenomics, although many factors such as concomitant medication, diet, substance abuse, smoking, age, gender, ethnicity, social support systems and illness play a role in the differences that are observed in treatment outcomes, the role that genetics plays is expected to be substantial [140]. This is highlighted by the fact that studies examining monozygotic twins have reported similarities with regards to treatment response [142], agranulocytosis [143] and weight gain [144]. Furthermore when examining siblings and first degree relatives, both treatment response and TD were reported to show heritable trends [145–147]. This demonstrated heritability in combination with the poor treatment outcomes that are observed, illuminates the value that pharmacogenomics has for the optimisation of antipsychotic treatments. If genetic tests can be developed to guide dosing prior to treatment, the determination of the type of drug and dose to use will require a once off test that will not change over time [87,148]. Thus, trial and error dosing, which can take months to years to optimise, can be replaced [120]. Figure 2.7 demonstrates the role that pharmacogenomics can play in the optimisation of antipsychotic treatment.

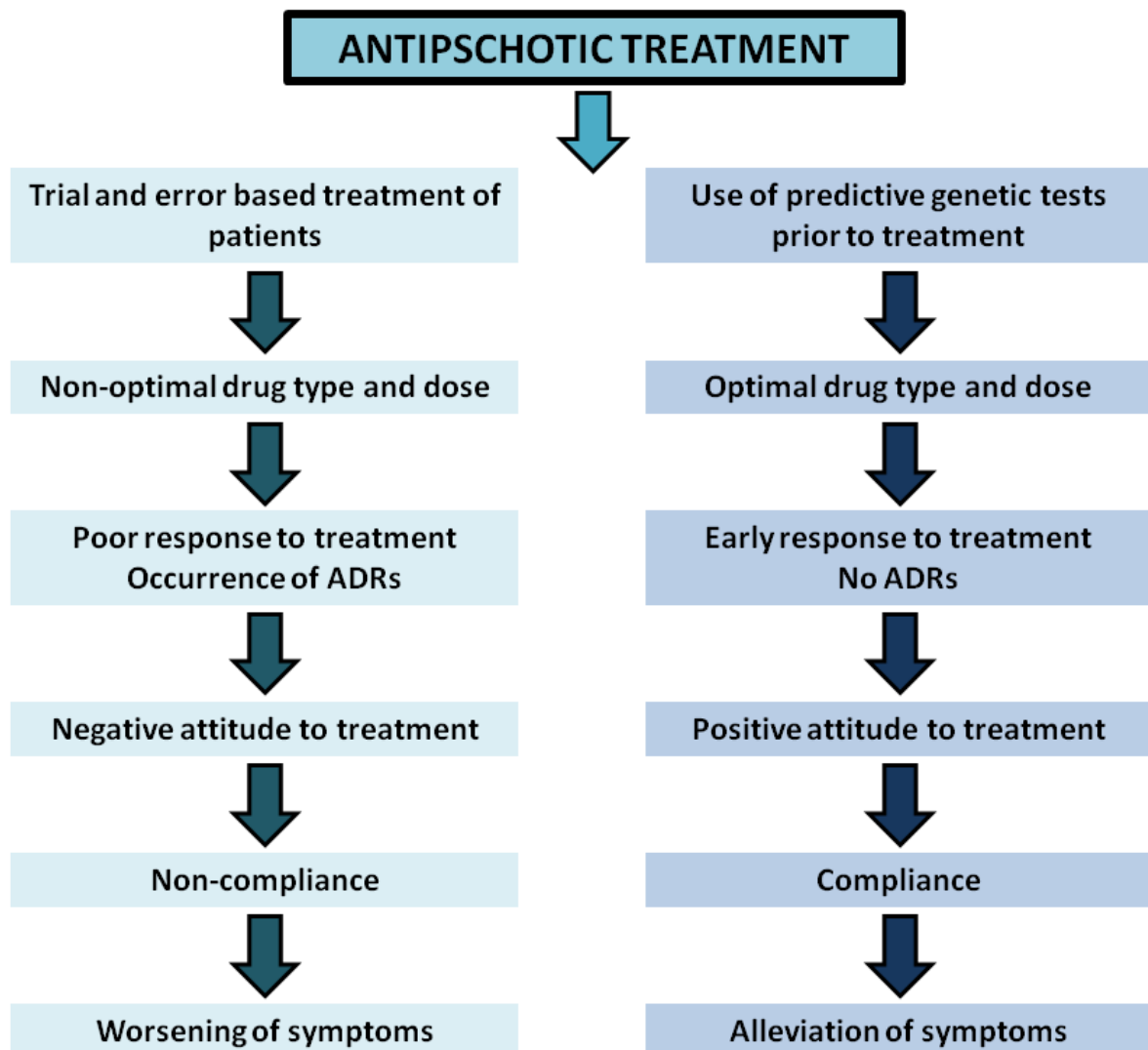


Figure 2.7: The role that pharmacogenomics can play in improving antipsychotic treatment outcomes.

The need for antipsychotic pharmacogenomics is highlighted by the availability of genetic tests that have been designed for antipsychotic pharmacogenetic applications (Table 2.1) [16,85,140,149]. The majority of these tests do not target specific ADRs, but examine variation in the cytochrome P450 (*CYP*) genes, which are responsible for metabolising antipsychotics. However, four tests have been described for weight gain related ADRs and clozapine-induced agranulocytosis. In terms of other success stories, six SNPs have been identified that are predictive of an individual's response to iloperidone. When examining individuals that showed a more than 20% improvement in PANSS scores, it was found that 75% of these individuals tested positive for at least one of the six SNPs [150]. The value of antipsychotic pharmacogenetics is further validated by the information provided on the PharmGKB database [151].

Table 2.1: Pharmacogenetic tests that have been designed for antipsychotic treatment

Test	Response outcomes	Reference
Psychotropic-Induced Metabolic Symptoms (PIMS) PhysioType™ System (Genomas)	Weight- and lipid-related ADRs	[152]
High-Low Metabolic Risk for Neuro-Psychiatric and Cardio-Metabolic Drugs (HILOmet) PhysioType™ System (Genomes)	Metabolic Syndrome	[153]
PGXPredict™: CLOZAPINE (Clinical Data Inc.)	Clozapine induced agranulocytosis	[154]
Clozapine Test (LGC)		[155]
Luminex Tag-It Mutation Detection Kit (Luminex)	<i>CYP2D6</i> and <i>CYP2C19</i> variants	[156]
AmpliChip® CYP450 Test (Roche) ¹		[157]
BRAINchip test (Progenika)		*
The GeneSightRx Psychotropic Test (AssureRx Health)		[158]
SureGene Test for Antipsychotic and Antidepressant Response (STA ² R) (SureGene; PGXL)		[159]
PHARMAchip (Progenika)	Treatment response outcomes	[160]
iPLEX® ADME PGx (Sequenom)		[161]
VeraCode® ADME CorePanel (Illumina)		[162]
Neurofarmagen® (AB-Biotics)		[163]
IMPACT genetic test (CAMH)		[164]

¹FDA approved; * Website no longer available.

Unfortunately, although tests are available for implementation in the clinic, physicians are often hesitant to put these tests into practice and they remain under-utilised. This applies not only to the lesser validated psychiatric tests, but also for well-validated examples such as the *HLA-B*5701* test for predicting abacavir hypersensitivity reactions, which has a sensitivity of 100% and a specificity of 96% [165,166]. The paucity of genotyping tests within the clinical setting may be related to a number of factors which include: (i) physician satisfaction with already existing monitoring processes; (ii) a lack of data pertaining to the cost-effectiveness of pharmacogenetic tests; (iii) impracticality of such tests (e.g. time period that is required to receive results); and (iv) a lack of conclusive results [167]. For this reason, more substantial evidence is required for pharmacogenetic results and this information needs to be combined with clear guidelines such as those proposed by the CPIC,

whereby peer-reviewed, evidence-based guidelines for gene/drug pairs are provided to facilitate the implementation of pharmacogenetics [168] (Table 1.1).

2.3.2 Candidate gene antipsychotic pharmacogenetic studies

Although not much is known regarding the mechanism of action of antipsychotics, certain pathways and genes have been identified that are known to be involved. When a drug is ingested, there are two main mechanisms of interest with regards to pharmacogenetic studies, namely the pharmacokinetic and pharmacodynamic processes. In simple terms, pharmacokinetic processes refer to what the body does to the drug, while pharmacodynamic processes relate to what the drug does to the body. Thus knowledge pertaining to the genes involved in these two processes guide the selection of the candidate genes that are examined [169].

The main genes involved in pharmacokinetics include those that contribute to the absorption, metabolism and transport of drugs. Due to the fact that these processes are important for many drug classes, the related genes are often classified under PharmGKB's Very Important Pharmacogenes (VIP) [170]. Although studies have examined many genes involved in pharmacokinetics, the main focus has been on the metabolism of the drugs, which occurs in two phases: phase I metabolism, which involves oxidation enzymes, and phase II metabolism, which involves conjugation enzymes such as N-acetyltransferases, thiopurine S-methyltransferases, UDP glucuronosyltransferases and glutathione S-transferases [171]. More specifically, studies have predominantly examined variation in the *CYP* genes which are involved in phase I metabolism [140]. These genes are good candidates for pharmacogenetic studies as they show a great deal of variation which affects the resulting activity of the enzymatic protein products.

To aid in the classification of the identified *CYP* variants and the effects that these variants have on metabolism, individuals can be divided into different metabolism groups based on the functionality of the genes present. These groups include poor metabolisers (PM) (no functional copies of the gene), intermediate metabolisers (IM) (only one functional copy of the gene, or two decreased function genes), extensive metabolisers (EM) (two functional copies of the gene) and ultra-rapid metabolisers (UM) (more than two functional copies of the gene) [140]. If, for example, an individual is a PM, the drug cannot be properly metabolised and there will be a build up of the drug, which may lead to ADRs. Conversely, if an individual is a UM, the drug will be metabolised too quickly and the desired response will not be achieved. In the case of prodrugs, the opposite effect is observed [172]. One of the best studied *CYP* genes is *CYP2D6*, which is involved in the metabolism of 25-30% of central nervous system drugs, including 40% of antipsychotics [73]. Additionally, this gene is highly variable with over 100 recorded alleles that change the function of the gene and include gene deletions, gene duplications, hybrid genes and gene conversions [173]. As is frequently observed with regards to pharmacogenetic alleles, the pattern of variation in this gene differs substantially between different population groups [174] (Figure 2.8), which could affect the treatment outcomes observed in these populations. By genotyping individuals prior to dosing and utilising the information regarding the functionality of the gene, the dose of the drug can be adjusted according to the metaboliser group that the genotyped individual falls into. As shown in Figure 2.9, PMs will require a lower dose of antipsychotic or antidepressants, while UMs will require a higher dose [175].



Figure 2.8: The differences in the frequencies of variants contributing to *CYP2D6* metabolising function across the globe [174]. Reprinted with permission from Wolters Kluwer Health.

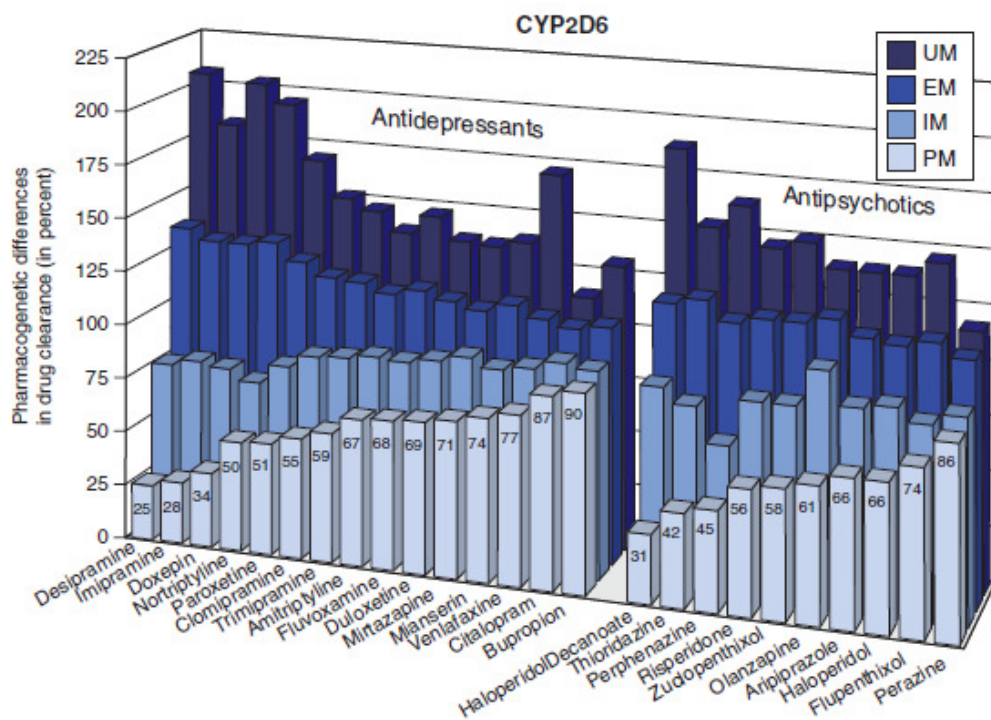


Figure 2.9: The suggested dosage adjustments for various antidepressants and antipsychotics based on *CYP2D6* metaboliser status. PM: Poor metaboliser, IM: Intermediate metaboliser, EM: Extensive metaboliser, UM: Ultra-rapid metaboliser [175]. Reprinted with permission from Nature Publishing Group.

The pharmacodynamics of antipsychotics refers to the mechanism of action of drugs, and includes genes such as those coding for the receptors that the drugs bind to in order to elicit a response. Variation in pharmacodynamic genes can affect both safety and efficacy pharmacogenetics; however, the genes that safety or efficacy studies focus on may differ depending on the phenotype that is being examined. With regards to efficacy pharmacogenetics, the main candidate genes relate to the receptor binding profiles of the drugs, and may also include genes involved in the pathogenesis of schizophrenia, such as *DISC1* and *DTNBP1* [73]. As mentioned previously all antipsychotics bind to dopamine receptors, while SGAs also bind to serotonin receptors. Therefore the genes encoding these receptors, as well as genes involved in these pathways, would be excellent candidates. As such, the most popular candidate genes include the dopamine receptors (*D1-D5*), transporters (*SLC6A3*) and metabolisers (*COMT*) as well as serotonin receptors (*HTR2A* and *HTR2C*), and transporters (*SLC6A4*) [172]. Furthermore, different antipsychotics bind with different affinities to the dopaminergic, serotonergic, noradrenergic, histaminergic and cholinergic receptors. Therefore genes in these pathways all make good candidates for antipsychotic efficacy pharmacogenetic studies. However, the selection of the candidate genes may be slightly different depending on the drugs under investigation [16].

With regards to safety pharmacogenetics, once again the candidate genes of interest may relate to the binding profiles of the drugs. However, knowledge based on the fact that FGAs are more likely to result in movement disorders and SGAs are more likely to result in weight gain ADRs can be used to guide the selection of candidate genes. The more frequent occurrence of movement related ADRs and hyperprolactinemia as a result of treatment with FGAs is related to the higher affinity that these drugs have for dopamine receptors. Therefore, genes relating to the dopamine pathway are good candidates for pharmacogenetic studies examining these ADRs. In a similar manner the higher affinity of SGAs for serotonin receptors and the corresponding greater likelihood of weight gain related ADRs, make genes in the serotonin pathway ideal candidates for weight gain related ADRs [176]. This is validated by the fact that serotonin plays a role in feeding behaviour and thus variants changing serotonin levels may explain the weight gain phenotypes that are observed [134,172]. In addition to the genes related to serotonin, other genes that may be involved in weight gain ADRs include genes involved in the regulation of appetite (e.g. the *SNAP-25*, *LEP*, *LEPR*, and histamine receptor genes); genes involved in lipolysis (e.g. the alpha adrenergic receptors); and genes involved in the metabolism of fatty acids and cholesterol (e.g. *INSIG2*) [134,176–178]. Thus, it is important to consider the biological underpinning of the ADR phenotypes when selecting candidate genes.

The narrow therapeutic window of antipsychotics [97,122] highlights the value of pharmacogenetics as a tool to optimise dosing. When examining review articles relating to “antipsychotic pharmacogenetics/genomics” that have been written within the last five years, it was noticed that the majority of research has focussed on candidate genes and that the most frequently referred to genes relate to both pharmacokinetic and pharmacodynamic processes, with an over-representation of genes related to metabolism processes as well as the dopamine and serotonin pathways (Table 2.2).

Table 2.2: The top 25 most frequently studied candidate antipsychotic genes as determined by relevant articles in the PubMed Database.

Gene	Function	Associated Pharmacogenetic Trait	References
<i>ABCB1</i>	Antipsychotic transporter	Treatment response, Weight gain	[16,120,127,171,176,177,179–182]
<i>CYP1A2</i>	Antipsychotic metabolism	Treatment response, Movement disorder	[15,16,119,120,171,176,177,179,183–186]
<i>CYP2D6</i>		Treatment response, Movement disorder, Weight gain, QT prolongation	[16,88,119–121,171,176,177,179–181,183–192]
<i>DRD1</i>		Treatment response	[176,193]
<i>DRD2</i>	Dopamine receptors	Treatment response, Movement disorder, Weight gain, Hyperprolactinemia, Neuroleptic malignant syndrome	[16,73,88,100,119–121,127,148,171,176–181,183–187,189–191,193–196]
<i>DRD3</i>		Treatment response, Movement disorder	[15,16,73,100,119–121,148,171,176,177,179,183–187,189–191,193,195]
<i>DRD4</i>		Treatment response, Movement disorder, Weight gain	[16,73,119,171,176,177,179–181,183,185,190,191,193,195]
<i>SLC6A3</i>		Dopamine transport	Treatment response
<i>COMT</i>	Dopamine metabolism	Treatment response, Movement disorder	[16,73,100,119–121,171,176,177,179,183,185,187,189,193,195–198]
<i>HTR1A</i>	Serotonin receptors	Treatment response	[16,73,100,148,171,176,179,189,195]
<i>HTR2A</i>		Treatment response, Movement disorder, Weight gain	[15,16,73,88,100,119–121,127,148,150,171,176,177,179–181,183–187,189–191,195]
<i>HTR2C</i>		Treatment response, Movement disorder, Weight gain	[15,16,73,88,119–121,127,171,176–181,183–189,191,195]
<i>HTR6</i>		Treatment response, Weight gain	[16,73,119,127,176,177,179–181,195]
<i>SLC6A4</i>		Serotonin transport	Treatment response, Movement disorder, Weight gain
<i>RGS4</i>	Inhibits <i>DRD2</i> and <i>DRD3</i> signalling	Treatment response	[16,73,171,176,177,179]

Gene	Function	Associated Pharmacogenetic Trait	References
<i>SNAP25</i>	Involved in the regulation of neurotransmitter release	Treatment response, Weight gain	[16,73,127,171,176,178–181,195]
<i>GNB3</i>	Integrates signals between receptors and effector proteins	Treatment response, Weight gain	[16,100,119,127,148,171,176,178–181,195]
<i>GSTM1</i>	Detoxification of drugs	Movement disorder	[16,171,183–185]
<i>SOD2</i>	Prevents damage caused by reactive oxygen species	Movement disorder	[121,148,171,176,183,185,189]
<i>LEP</i>	Plays a major role in the regulation of body weight	Weight gain	[16,121,127,171,176,178–181,184,186,189]
<i>LEPR</i>	Leptin receptor	Weight gain	[16,121,127,176,180,181,184]
<i>MTHFR</i>	Folate and homocysteine metabolism	Treatment response, Weight gain	[16,176,177,179–181,195]
<i>HLA-DQB1</i>	Involved in immune response	Agranulocytosis	[16,121,136,176,186,189]
<i>TNF</i>		Treatment response, Weight gain, Agranulocytosis	[16,121,127,136,171,176,178–180]
<i>BDNF</i>	Member of the nerve growth factor family	Treatment response, Movement disorder, Weight gain	[16,73,127,171,176,179–181,183,195,199]

2.3.3 Genome-wide association studies for antipsychotic pharmacogenomics

Many candidate genes that may be involved in antipsychotic treatment outcomes have been identified. Although the role that these genes play in antipsychotic response are biologically validated, not many of the findings have been consistently replicated. The lack of replication may either be related to associations detected as a result of false positives, or the fact that studies are very heterogeneous in nature with reference to genotyping strategies, classification of phenotypes and the ethnicity of patients [172,195]. Nonetheless part of the inconsistency in these results may be related to gaps in our knowledge regarding the mechanism of action of antipsychotics, an inability of these studies to examine all genes and variants simultaneously and limited power to detect associations. It is for this reason that GWAS are expected to identify additional genes that are involved in pharmacogenetic phenotypes. However, although there are a multitude of antipsychotic pharmacogenetic studies examining candidate genes, only a handful of studies have used GWAS approaches [53].

The first antipsychotic GWAS to be published examined antipsychotic induced weight gain in a small sample of individuals from 21 families and identified a region on chromosome 12, which contained the *PMCH* gene [200]. In a second GWAS, which used 100 individuals to detect an association with TD, the GABA pathway was identified as a potential contributor to the development of TD [201]. Since then, just over ten GWAS have been published that meet the requirements for appearing on the HuGE Navigator GWAS Integrator [202] (Table 2.3). These GWAS have examined several different aspects related to the pharmacogenomics of antipsychotics, ranging from treatment response phenotypes to ADRs such as TD and weight gain. Unfortunately the sample sizes used in these studies remain small, with the largest cohort obtained from the CATIE study (n=738). In addition, the majority of the published GWAS have used this CATIE cohort, which is a heterogeneous cohort of patients receiving various different antipsychotics. Thus, this study is not ideally suited to pharmacogenomic research. Although the majority of the *P*-values obtained from the published GWAS do not meet genome-wide significance, there is one study that obtained a *P*-value a few orders above this threshold. The success of this study may be related to the fact that this study was designed with pharmacogenomics in mind. The study utilised a longitudinal design, which made use of well characterised patients, as well as replication cohorts. The patients were drug-naive prior to treatment and drug plasma levels were used to assess compliance, thus more accurate assessment of phenotypes was possible [203]. This highlights the importance of careful study design.

Table 2.3. The GWAS and top variants significantly associated with antipsychotic pharmacogenetic traits as reported on the HuGE Navigator GWAS Integrator [202].

Variant	Published Gene	Disease/Trait	Reference	Discovery Cohort Size	Replication Cohort Size	P Value
rs489693	<i>MC4R</i>	Weight gain	[203]	139	205	6×10^{-12}
rs1405687	Intergenic	Treatment response	[204]*	738	NA	5×10^{-8}
rs7838490	<i>MMP16</i>	BMI and cholesterol	[205]	594	NA	6×10^{-8}
rs7669317	Intergenic	EPS	[206]*	738	NA	8×10^{-8}
rs11240594	<i>SLC26A9</i>	Treatment response	[207]*	738	NA	1×10^{-7}
rs888219	Intergenic	Treatment response	[208]*	738	NA	2×10^{-7}
rs4959235	<i>SLC22A23</i>	QT interval prolongation	[209]*	738	NA	2×10^{-7}
rs11110077	<i>ANKS1B</i>	Working memory	[207]*	738	NA	4×10^{-7}
rs17706989	<i>WWOX</i>	Reasoning	[207]*	738	NA	1×10^{-6}
rs993648	<i>CERKL</i>	QT prolongation	[210]	183	NA	3×10^{-6}
rs12476047	<i>FIGN</i>	Treatment response	[211]*	397	NA	3×10^{-6}
rs10805321	Intergenic	EPS	[212]	409	NA	9×10^{-6}
rs3943552	<i>GLI2</i>	TD	[213]*	327	170	6×10^{-5}
rs11851892	<i>NPAS3</i>	PANSS-T score	[214]	106	104	9×10^{-5}

*These studies were performed using samples obtained from the CATIE study. BMI: Body Mass Index; EPS: Extrapyramidal Symptoms, TD: Tardive Dyskinesia, PANSS-T: Positive and Negative Symptom Scale-Total score.

This being said, it is often difficult to obtain large and well characterised cohorts. Although the effect sizes attributed to pharmacogenetic traits are expected to be larger than those attributed to diseases [26], large cohorts of more than 1000 individuals are often required in order to reach genome-wide significance. Unfortunately, it is difficult to obtain large cohorts of similarly treated schizophrenia patients, which exhibit the pharmacogenetic traits of interest [176,215]. Highlighting these issues, it has been reported that 40% of genetic studies examining antipsychotic response have cohorts smaller than 100 and 20% have cohorts smaller than 50 [189]. Furthermore, although replication is vital for the validation of GWAS results, it may be very difficult to obtain cohorts of patients that have been treated with the same drug in a similar manner and whose treatment outcomes have been assessed using comparable measurements [27,215]. Nonetheless, as we move forward into the era of genomic research, study design is of utmost importance and many factors need to be taken into consideration. These factors include accurate diagnoses and measurements of response, as well as the use of homogeneous cohorts and the assessment of confounders.

With regards to the diagnosis of schizophrenia, patients should receive a DSM, ICD or equivalent diagnosis of schizophrenia before being included in studies examining schizophrenia. It has been calculated that a reduced diagnostic accuracy (i.e. misdiagnosis) can seriously impact on the number of samples required to detect an association. In other words, smaller cohorts of better characterised patients will have the same power as larger cohorts of less well characterised patients [216]. In order to successfully study treatment response outcomes in schizophrenia, the following cohorts would be ideal. Patients who are experiencing their first episode of schizophrenia (therefore illness course can be eliminated

as a confounder) and are not affected by disease comorbidity should be recruited. These patients should be previously drug-naïve (therefore treatment duration will be eliminated as a confounder) and once they begin treatment they should receive the same drug (to eliminate differences in drugs), which they receive as an injectable (to eliminate non-adherence as a confounder), without concomitant medication. These cohorts of patients should be examined for as long as possible so as to examine the effects of medication over a long time period and patients should be regularly assessed at the same time periods [117,176,177,180,195,217].

With particular reference to the study of treatment response phenotypes, it is important to utilise patients that exhibit a range of treatment response phenotypes. When using a cross-sectional cohort of multi-episode patients, it remains likely that those patients who have responded well to medication will be filtered out of the examined cohort and thus the sample will be biased towards non-responders, making association analyses difficult. [113]. The value of the use of FES cohorts in the assessment of treatment response was demonstrated by Zhang *et al.* [218], where it was reported that a 50% greater effect size was observed with regards to the -141C *DRD2* variant and treatment response when utilising a meta-analysis in FES cohorts rather than multi-episode cohorts. This diluting of samples may also be attributed to the fact that 20-72% of chronic schizophrenia patients are non-adherent to medication. To demonstrate the importance of adherence in pharmacogenetic studies, a sample of 250 patients with 90% adherence will be better powered than a sample of 400 patients with 50% adherence [218].

Once these well characterised cohorts have been collected, it is important that treatment outcomes are calculated using standardised measurements. However, unfortunately unlike other disorders where biochemical/physiological tests are available to diagnose symptoms, the assessment of the symptoms present in schizophrenia is reliant on the use of various scales that are given scores based on observation. Thus, these measurements remain susceptible to bias due to the subjective nature of these scales. Furthermore, the diagnostic categories are not clearly delineated from one another and they are often shared with other psychotic disorders, further complicating the accurate assessment of phenotypes [42]. Nonetheless, careful and regular assessment with the relevant scales may help to alleviate these issues. The most well validated scales for these assessments include the PANSS or SANS and SAPS scores for treatment response measurements; the abnormal involuntary movement scale, extrapyramidal symptom rating scale, Simpson-Angus scale, Barnes Akathisia Rating Scale or Unified Parkinson's Disease Rating Scale for EPS; and body mass index, hip circumference, waist:hip ratio, lipid status and blood pressure for weight gain related ADRs [85]. Using standardised scales will help to ensure that measurements are comparable over time and across patients. In addition, different symptom domains should be considered separately and outcome measurements should be compared to baseline measurements, further highlighting the importance of recruiting patients at their first episode of schizophrenia [217].

While current schizophrenia GWAS have utilised strategies that rely on scanning common genetic variants in large cohorts of schizophrenia patients, it appears that antipsychotic pharmacogenomic studies may require better characterised cohorts of patients and technologies that are able to examine the entire spectrum of variation. By utilising WGES of

well characterised patients, the likelihood of identifying causal variants may be increased due to the ability of these technologies to detect both rare and common variants within functional regions. Thus, results obtained from these WGES studies are more likely to have direct relevance in the clinical setting. Unfortunately, to date there are no studies utilising WGES approaches to uncover the genetic factors relating to antipsychotic treatment outcomes.

2.4 The study of antipsychotic pharmacogenomics in the South African context

2.4.1 Resource limited settings and unique cultural considerations

In LMIC, there is a high rate of poverty, inequality and violence, which may exacerbate the onset of schizophrenia [219]. In South Africa, these factors have been heightened by the racial discrimination and political violence of the apartheid regime [220]. The lack of education, financial resources, accessibility to health care services, high rate of co-morbid disease and substance abuse may all act as stressors and may prevent the effective treatment of patients [221]. In the Western Cape of South Africa, the use of methamphetamine (Tik) is very high and this plays a role in heightening the already frequent occurrence of psychiatric disorders, as well as complicating the treatment thereof [222]. The lack of resources in South Africa is highlighted by the fact that there is only one psychiatrist per 100 000 individuals. Most of these psychiatrists work in Cape Town and Gauteng (Johannesburg and Pretoria regions), with only 4.7% of psychiatrists working in state hospitals in rural areas [223]. This emphasises the disparities in access to health in an already under-resourced country. In addition, the overburdened health care systems of South Africa appeared to experience a decline in resources from 1997 to 2005. Per 100 000 individuals, the number of beds declined from 48 to 28, the number of mental health staff decreased from 19.5 to 11.95, the number of psychiatrists declined from 0.4 to 0.3 and the number of nurses declined from 15.6 to 10.1 [4]. Furthermore, it has been reported that only a quarter of South Africans with mental disorders sought out treatment and, perhaps even more shockingly, only 5.7% received formal mental health care [224]. This lack of treatment may be related to a lack of resources, as well as the high level of stigma associated with schizophrenia, which is reportedly higher than in other mental disorders [225]. A study examining the stigma associated with South African Xhosa schizophrenia patients reported that 52% and 44% of the family members believed schizophrenia patients to be dirty or dangerous [226]. Other comments included the fact that they were unpredictable, tense, worthless, delicate, slow, weak and foolish. Another South African study reported that 65% of patients felt as though they had been discriminated against as a result of their illness. In addition, many patients had experienced some kind of abuse, with 58% of the patients experiencing name-calling, 58% experiencing verbal abuse and 39% experiencing physical abuse. The abuse was more frequent among *isiXhosa* speaking individuals, when compared to English and Afrikaans speaking patients [227].

In addition to this abuse, treatment of patients is often limited, which may be accounted for by cultural beliefs. In a study examining Xhosa schizophrenia patients, 67% of the family members believed that the symptoms of the disorder could be attributed to *amafufunyana*, which refers to witchcraft or spirit possession; or *ukuthwasa*, which refers to a calling to

become a traditional healer. *Amafufunyana* has been associated with undressing, aggressive behaviour and psychomotor agitation. *Ukuthwasa* has been associated with social withdrawal, irritation, restlessness, screaming and shouting, dancing and auditory hallucinations [228]. In the context of South Africa, a large percentage of patients will use traditional healers as their first line of treatment, before seeking medical advice [229], with an estimated 80% of the South African populations visiting the 200 000- 350 000 traditional healers in the country [230]. In the context of schizophrenia, when examining a cohort of predominantly Zulu patients, it was observed that half of the patients believed that their illness was as a result of a spiritual cause and 38.5% of these patients consulted a traditional healer [231]. Another study reported that 80% of Xhosa schizophrenia patients consulted a traditional healer and of these 80%, 53% were diagnosed with *amafufunyana* (50.9% of which had a family history of schizophrenia) and 4.5% were diagnosed with *ukuthwasa* (100% of which had a family history of schizophrenia) [228]. Of further concern, patients who are diagnosed by traditional healers experience a longer DUP [231]. As a longer DUP has been associated with worse treatment outcomes [232], this may have implications for the well-being of these patients. Thus, it may be important to form a bridge between western medicine and traditional healers.

The inefficient treatment of schizophrenia places a heavy financial burden on South African health services. In addition to cultural sensitivities, these poor treatment outcomes are influenced by a lack of guidelines for the treatment of psychiatric disorders and differences in ethnicities between patients and clinicians, making communication and interpretation of symptoms difficult [233]. It is important that language barriers are overcome and that information is provided in the home language of the patient in a way that is sensitive on both a cultural and social level [230]. Furthermore, due to less regular check-ups accompanied with an inability to regularly monitor patients, drugs such as clozapine which are associated with dangerous ADRs, are less likely to be prescribed even though they have been documented to show better treatment outcomes in treatment resistant patients [233]. Even if patients are able to gain access to treatment, non-adherence is of serious concern and is influenced by the difficulty in obtaining antipsychotics. Often patients need to queue outside the clinic from 4 am. They may have to wait for hours in order to obtain medication and stand the risk of being mugged during this time. Many patients choose to use half of the prescribed dose so that they can make less frequent visits to the clinic [234].

Of further concern, treatment regimes and guidelines available to high income countries are often not applicable in the context of LMIC. When examining the use of antipsychotics in the treatment of schizophrenia, the costs of these treatments need to be considered. In this regard several studies, including the CATIE and CUtLASS studies, have reported that costs associated with the use of FGAs are lower than those of SGAs [235]. For this reason, studies that have examined the cost-effectiveness of treatment recommend the use of FGAs in combination with psychosocial interventions to reduce the costs associated with the disorder, while keeping the costs associated with the treatment of the disorder to a minimum [236,237]. Highlighting the high cost associated with SGAs, it has been reported that in the last decade the global expenditure on antipsychotics in the USA has increased approximately 20 fold, from \$0.5 billion per year to \$15 billion per year. This is largely attributed to the more frequent use of SGAs [185], which are approximately ten times as expensive as FGAs [86]. Nonetheless, although the efficacy of FGAs and SGAs has been

reported to be very similar, it is generally recommended that SGAs are given as first line treatment as they are less likely to result in EPS. Unfortunately, in Africa, where resources are limited, the cheaper FGAs are usually utilised. More specifically, in South Africa only schizophrenia patients who have access to private psychiatric facilities (approximately 30% of the population) have access to SGAs. Thus, the majority of South Africans will need to use FGAs as their first line of treatment [86]. Therefore, studies examining antipsychotic pharmacogenomics in the African context should focus on FGAs. If pharmacogenomic interventions to alleviate the ADRs associated with FGAs can be developed, the cost-effectiveness of the treatment strategies can be further improved [140].

Interestingly, even though access and quality of treatment in less economically developed countries is limited, it has been reported that schizophrenia patients in these regions have been shown to respond better to treatment than those living in higher income countries such as Europe. This has been observed both with respect to response and remission rates [238]. It has been suggested that the African collective culture provides a greater level of support, when compared to the Western culture of individualism, and for this reason treatment outcomes in these patients are better [239]. However, the validity of these results has been questioned. The findings may be influenced by the fact that there may be a higher rate of misdiagnosis in developing countries, which may skew the results, and not all patients who are suffering from schizophrenia in the developing world may have access to the hospitals sites where these studies were performed [221]. Nonetheless, although there are many hurdles in examining schizophrenia and the treatment thereof in LMIC, South Africa remains a leader in Africa with regards to mental health services and research [2]. Egypt, Nigeria and South Africa provide the largest number of biomedical research papers in Africa [18] and South Africa is considered one of the top five LMIC with regards to psychiatric research [240]. Thus, the future of research into the optimisation of antipsychotic treatment in South Africa is bright.

2.4.2 The rainbow nation and implications for genomic studies

Archbishop Desmond Tutu aptly named South Africa the rainbow nation due to the many diverse individuals that make the country their home, which is reflected by the eleven official languages spoken in South Africa. The populations residing within South Africa range from homogenous (Afrikaner population) to admixed (Coloured population) to ancient and genetically diverse (African populations) (refer to Warnich *et al.* [230] for more details regarding the history and genetic composition of the South African populations). These populations offer unique challenges, as well as advantages for genomic research. The high diversity and low LD observed in the African populations, although not ideal for GWAS, are valuable for WGES and fine mapping strategies (refer to Chapter 3 for more details) [241]. In contrast the homogenous Afrikaner population is less suited for fine mapping, but is well suited to GWAS due to the ability to tag large areas of the genome with relatively few variants. Lastly, the admixed nature of the South African Coloured (SAC) (referred to as such by the National Census) or South African Mixed Ancestry population allows for the opportunity to identify pharmacogenomic variants of relevance to several different populations [242].

The admixture present in the SAC population, along with the differences in genetic make-up observed between the different South African populations, brings to light a very important consideration with regards to the implementation of pharmacogenetics in South Africa. It seems unlikely that a uniform panel of variants will be identified that will apply to all individuals residing within the country and thus an individualised approach to medicine may be required. Furthermore, it should be noted that although South Africa provides rich sources of populations with unique advantages for pharmacogenomic research, the well-being of these individuals should be considered before research is performed. It is becoming more popular for researchers from developed countries to perform studies on local South African populations and for this reason the concept of genomic sovereignty is important. This ensures that genomic material that is utilised from South African individuals is protected and that the generated research is beneficial to the local individuals from whom the material was sourced [230].

To date, the only published studies relating to the genetics of schizophrenia in South Africa have been performed in the Xhosa and Afrikaner populations [240]. Currently, there are no published studies that have utilised the SAC to identify associations with schizophrenia. This may be related to the fact that the high admixture in this population complicates genetic association studies. In order to correct for this admixture, ancestry informative markers (AIMs) are required. These AIMs should be selected based on the different frequencies at which these markers occur in the populations that have contributed to the ancestry of the SAC population [243]. By using these AIMs to perform principle component analyses, the principle component scores can be used as covariates in the statistical analyses in order to correct for false positives that may be caused by population stratification. Thus, the development of AIMs that have been designed specifically for the highly admixed SAC population are essential in order for association studies to be performed in these populations.

With regards to antipsychotic pharmacogenetic studies, more than half of these studies have been performed in European-descent individuals, with the majority of the remaining studies performed in Asian individuals [244]. This leaves African populations poorly represented. In addition, although large resequencing projects such as the 1000 Genomes Project have undertaken to characterise the human variome, the diverse populations present in South Africa remain under-represented (refer to Chapter 3 for more details) [241]. Therefore, it is important that pharmacogenomic studies are performed in South African populations and that the variation present in these populations is characterised. With reference to antipsychotic treatment, it has been reported that South African individuals of different ethnicities respond differently to antipsychotic treatment, such that SAC individuals respond best (32% response rate), followed by Africans (24% response rate) and European descent individuals (9% response rate). It has also been reported that Asians require a lower dose of antipsychotics to achieve a treatment response [245]. The contribution of Asian ancestry to the SAC group may partially explain the better response rates in this group of individuals. These same ethnic differences were reported for the development of TD and antipsychotic induced weight gain, with African ancestry individuals reportedly more likely to develop TD and experience weight gain after antipsychotic treatment than European descent individuals [246–248]. There are many factors that could explain the differences in response rates that are observed between individuals of different

ethnicities. These include diet, body mass, substance abuse, and importantly in the context of this study, genetics [245]. The effect of genetic variations on population specific response to medication has been demonstrated with regards to the development of Steven Johnson syndrome (SJS) as a result of carbamazepine treatment. It has been observed that individuals that carry the *HLA-B*502* allele are likely to develop SJS, however this is only the case for Asian individuals [249]. Due to the fact that large differences in pharmacogenetic allele frequencies are observed between different South African populations [240], it is important to determine how these and other variants contribute to the differences that are observed with regards to treatment outcomes. Thus the use of unique South African populations in antipsychotic pharmacogenomic sequencing studies may provide much needed answers for the optimisation of the treatment of schizophrenia patients.

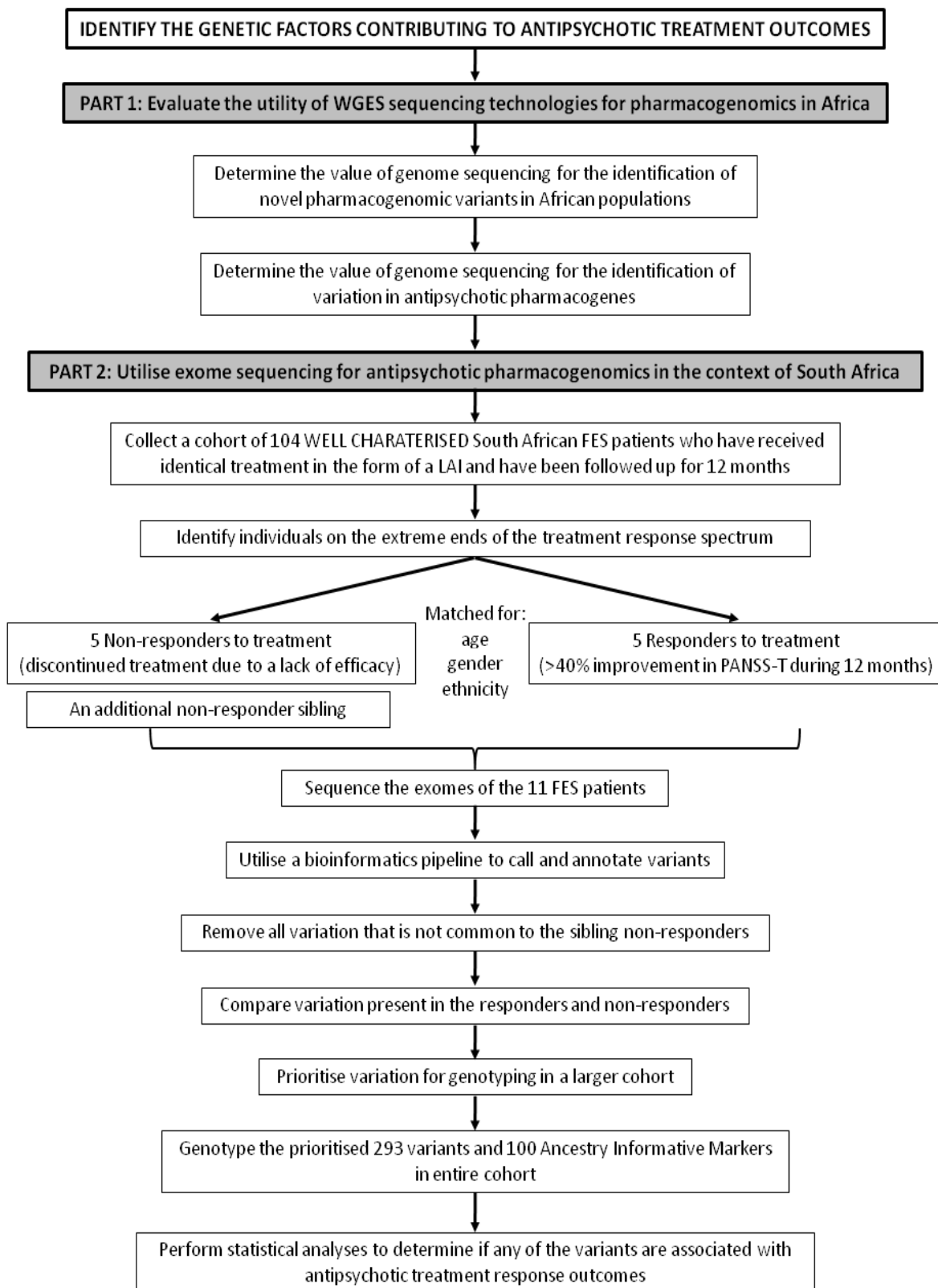
2.5 Aims and objectives

This project aims to identify genetic variants contributing to antipsychotic treatment response outcomes.

This will be achieved through the completion of the following objectives:

- 1) Evaluate the utility of WGES for antipsychotic pharmacogenomics in African populations
- 2) Identify and characterise the coding variation present in a subset of FES antipsychotic non-responders and responders with the use of exome sequencing
- 3) Compare the patterns of variation occurring in antipsychotic non-responders and responders
- 4) Identify genetic variants associated with antipsychotic treatment response in the South African FES cohort

2.6 Study outline



**PART 1: Evaluation of the utility of whole genome and
exome sequencing for antipsychotic
pharmacogenomics in Africa**

CHAPTER 3: Whole genome resequencing in pharmacogenomics: moving away from past disparities to globally representative applications

Drögemöller Britt I¹, Wright Galen EB¹, Niehaus Dana JH², Emsley Robin A², Warnich Louise¹

¹Department of Genetics, Stellenbosch University, South Africa

²Department of Psychiatry, Stikland Hospital, Stellenbosch University, South Africa

Reprinted with permission from Future Medicine.

3.1 Summary

Africa suffers from a high burden of disease; nonetheless, it has been one of the most under-represented continents with regard to genomic research. It can be argued that this disproportionate research is related to the fact that the genome architecture of African individuals is poorly suited to SNP-based genome-wide association studies, given existing genotyping platforms. However, this argument is no longer plausible with the arrival of next generation sequencing technologies, which allow for the analysis of entire genomes. Using pharmacogenes to critically examine the merit of next generation sequencing technologies in pharmacogenomics, we found a substantial amount of novel/uncharacterised variation, which was predicted to alter protein function. This variation was predominantly observed in African individuals, emphasising the benefit of next generation sequencing technologies specifically for these individuals. We also observed an improvement in the reliability of sequencing technologies in a relatively short time. Therefore, as sequencing technologies develop and decrease in cost, the ability to reliably detect variation will improve and these technologies will begin to replace other less comprehensive genotyping assays.

Keywords:

Africa, next generation sequencing, pharmacogenomics, pharmacogenes, genomics

3.2 Inequalities in genomic research

There has been a large disparity in genomic research, with approximately 75% of GWAS performed in European descent individuals [250], while the remaining GWAS provide a poor representation of African individuals. When consulting the existing database of GWAS literature [202] (accessed 28th March 2011), out of a total of 5 854 possible “GWAS hits”, 465 were related to the search term “African”. However, from these results, only six studies referred to specific populations residing within the continent (i.e. West Africans, Nigerians, Malawians, Gambians and Ghanaians [251–256]), while all other studies referred to African Americans or African descent individuals. As can be seen in Figure 3.1, the “GWAS hits” that have been obtained from African populations represent a very small portion of the continent and only fall in areas of the Niger-Kordofanian language family distribution [257,258]. Since Africa is host to 30.5% of the world’s 6 909 living languages [259], it is not realistic to refer to Africans as one homogenous population group, nor is it accurate to infer the results obtained from the highly admixed African American GWAS onto present day African populations [260].

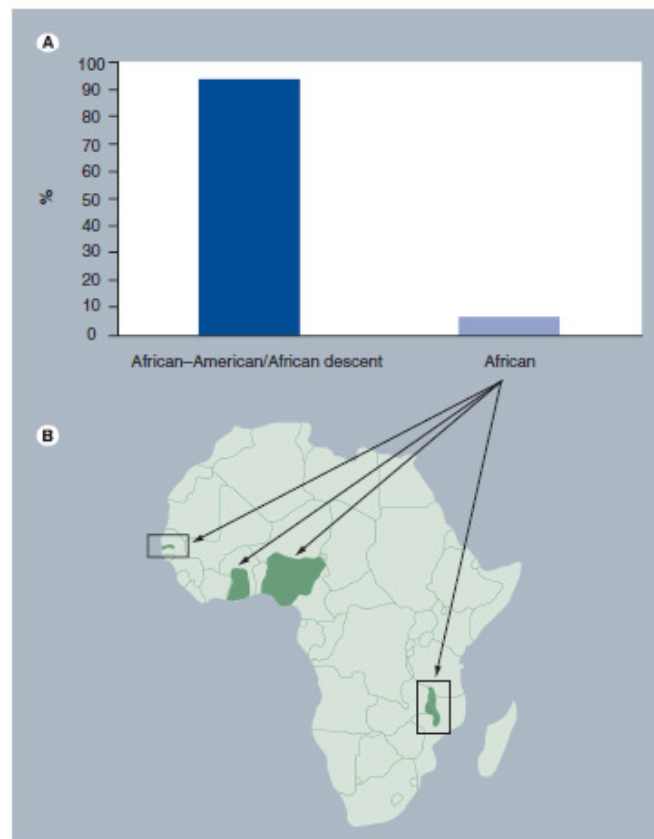


Figure 3.1 A: Percentage African “GWAS hits” [202], which represent ~8% of the GWAS performed to date, stratified by those studies using specific populations within Africa and those using populations defined as African American or African descent. **B:** The countries in Africa from which the GWAS cohorts were gathered. This image highlights the fact that those few studies that have used populations from within Africa represent only a very small portion of the continent

The under-representation of studies examining populations residing within Africa seems ironic when considering that out of the 28 countries worldwide that have a death rate higher than 1 850 per 100 000, 26 of these countries are situated in Africa [261], making Africa one of the continents most likely to benefit from the translation of GWAS results. Furthermore, due to the shortage of genomic research in Africa, those GWAS that have been performed include the more frequently studied TB, HIV and malaria epidemics [252,255,256], while other disorders, such as complex psychiatric disorders, are absent. The lack of psychiatric GWAS in Africa is further highlighted by the absence of these studies in the PGC, whose goal is to conduct meta-analyses utilising GWAS data for psychiatric disorders [60]. This absence is a result of the better genome wide coverage and greater amounts of data available for European descent individuals; however, it is the long term goal of the PGC to include African descent individuals [60]. This emphasises a need for additional studies within Africa that move beyond the more frequently examined epidemics to cover a wider variety of diseases and disorders. It should also be noted that one of the predominant explanations for the lack of genomic research in Africa can be attributed to a lack of funding, resources and infrastructure within the continent. This in turn results in a lack of expertise and subsequently a shortage of comprehensive genetic/genomic studies pertaining to African individuals [249]. Without access to large cohorts of well characterised

African individuals, current technologies have been unable to incorporate the genomic architecture of these populations into genome wide assays.

Fortunately the realisation that we need to address these research disparities has come at the right time. Although the GWAS of the past have focussed on European descent populations, it can be argued that these types of GWAS, utilising SNP genotyping, are well suited to the genomic architecture of European descent individuals, while the low LD present in African individuals is uniquely suited for fine-mapping [250,262]. Therefore by utilising individuals of African Diaspora in resequencing studies, we may be able to more easily elucidate causal variants, owing to the lower LD in these individuals. By identifying causal variants as opposed to those that are merely in LD with these variants, we may be able to increase the amount of evidence-based guidelines to aid in the application of genomic results into the clinical setting [263]. Furthermore, Africa provides the ultimate setting for the translation of genomic results, due to its high disease burden and ample disease cohorts [264,265], with reference to both communicable [266,267] and non-communicable diseases [268,269]. However, before Africa enters the genomic research arena, it is important that researchers focus on utilising techniques that will benefit from the complex architecture of African genomes. Bearing this in mind, there is evidence to suggest that SNP-based GWAS employing commercial microarrays may be replaced, in many areas, with the use of NGS technologies [270]. Therefore, Africa may be entering the field of genomics at the perfect time, coinciding with the explosion in sequencing technologies.

3.3 Next generation sequencing technologies

This year, as we commemorate the decade that has passed since the human genome sequence was published, a retrospective look at sequencing reveals one of the most rapidly advancing technologies to date. Although Sanger sequencing, published in 1977 [271], earned Frederick Sanger the Nobel Prize in 1980; the sequencing of the human genome using this technique took thirteen years to complete [272]. Today, with the use of NGS technologies, the whole genome sequence can be obtained in a few weeks and the associated cost has dropped about a million fold [273], making the field of genomics one of the leaders in science and technology at present. Even so, the potential of high throughput sequencing remains largely unharnessed and there is much room for improvement. By smoothing out the associated flaws and implementing WGES on a regular basis, it is hoped that the criticism associated with the lack of translatable results obtained from the Human Genome Project (HGP) can be addressed and genomic applications in the clinical setting will become routine.

Exciting as this technology may be, it is important that the biases of past genomic research are not repeated and that all populations are equally represented. Even though European ancestry individuals have been over-represented in GWAS to date, the lack of clinically applicable results obtained from these studies have fortunately prevented a large increase in global health disparities as a result of this research bias. However, if this uneven research continues into WGES studies, which are more likely to identify causal variants, the knowledge gap and health disparities between developing and developed countries are likely to increase, to the detriment of developing countries [249]. Unfortunately, even in the short time that WGES technologies have been available, research disparities are already emerging. These disparities refer both to sequencing-related equipment and resequencing data. While a substantial amount of both Caucasian and Asian genomes are being

sequenced, the resequencing of African genomes is lagging behind [274]. Due to the fact that all humans originate from Africa and African populations have largely avoided the bottleneck effects experienced by non-Africans [275,276], genomic characterisation of these populations should provide the most comprehensive catalogue of human variation. Therefore, if anything, these individuals should be over-represented rather than under-represented in genomic research. As mentioned in the previous section, the lack of African GWAS data can be justified by the difficulty in completely capturing the variation in African genomes using Eurocentric SNP-based genotyping techniques, due to high levels of genetic variation and low levels of LD [250]. However, WGES technologies allow for the almost complete capture of high levels of genomic variation, and low LD may even improve the ability to identify causal variants. Unfortunately, although the sequencing technologies for genomic studies are available and affordable, and Africa provides the required disease cohorts for these studies [264,265], WGES in African individuals remains limited.

These issues are, however, being addressed and the 1000 Genomes Project is in the process of resequencing 100 individuals from each of the five selected African populations from Nigeria, Kenya, Gambia and Malawi, as well as an additional population from either Sierra Leone or Nigeria [277]. These data should play a role in adding to the knowledge of not only the African genome, but also the human variome. This is important, as examination of the available resequencing data has demonstrated that African genomes consistently exhibit more variation, both novel and known, than non-Africans [278–281]. Unfortunately, with the exception of Kenya and Sierra Leone, the populations that are included in the 1000 Genomes Project originate from the same areas in Africa where the “GWAS hits” of the past have been obtained [251–256] (Figure 3.1). Therefore, once again the majority of Africa remains unaccounted for. However, every sequenced genome will add to our understanding of human variation and will create a stepping stone for future resequencing projects in more African populations. Through the careful implementation of WGES in Africa, taking into consideration the possibility of identifying causal variants contributing to diseases, it is hoped that genomic findings can be translated into the clinical context to improve health. These results are likely to have the biggest impact in Africa where improvement of health care is urgently required.

3.4 Pharmacogenomics: A front runner for translation of genomic results into the clinical setting?

Although the HGP was expected to play a role in understanding disease, the most tangible results were in fact related to personalised medicine and the effective treatment of disease after diagnosis [282]. More specifically, pharmacogenetics/genomics was both predicted and has been proven to show promise for clinical applications [19,283]. This is demonstrated by the translation of results from pharmacogenetic studies into the clinical setting [284,285], in combination with the number of FDA-approved drugs with pharmacogenomic information on their labels [22]. Furthermore, in some cases the cost-effectiveness has been calculated with regard to the utilisation of specific pharmacogenetic tests to improve drug efficacy and decrease ADRs [286–292]. Although these studies have shown that the pharmacogenetic tests are in some cases very expensive, which is a particular hindrance for implementation in developing countries, they have also shown that pharmacogenetic tests can be cost effective if the appropriate considerations are made and the allele frequencies of the specific variants are high enough in the population being tested

[287–291]. Results such as these, providing evidence for cost-benefit outcomes, are important as it has been widely reported that ADRs and treatment failure are responsible for large economic and healthcare burdens worldwide [293–295]. Additionally, these results emphasise a need to characterise and understand the variation present in the genes influencing drug response and to determine the role they play in treatment outcomes in different populations.

Unfortunately, studies providing examples of the cost effectiveness of pharmacogenomics within Africa are missing, which may in part be attributed to a lack of data pertaining to the specific variants affecting treatment outcome. This is of serious consequence as Africa is suspected to be burdened with a high rate of non-optimal treatment regimes. One of the reasons that Africa is burdened by ADRs and poor drug efficacy is the high prevalence of HIV/AIDS, which requires lifelong treatment with antiretrovirals [296]. Not only are patients more likely to experience ADRs due to the presence of the disease, but the use of concomitant drugs increases the likelihood of drug-drug interactions, which subsequently results in non-optimal treatment outcomes [294,297]. Furthermore, diseases and the treatment thereof are aggravated by poverty and a lack of resources, and as mentioned previously in Chapter 1, although 10% of the global burden of disease can be attributed to diseases occurring in poor countries, only 1% of new drugs are developed for the treatment of these diseases [298]. Therefore, not only are the relevant drugs not developed for Africa, but those drugs that are developed are predominantly designed using European descent individuals as a reference, which is unlikely to be an optimal fit for the highly diverse genomes of individuals residing within Africa. Therefore, research to determine the presence and effect of variation in pharmacogenes (i.e. pharmacogenetic/genomically relevant genes) needs to be performed in Africa to assess the types and dosages of medication required for optimal treatment.

Validation for the requirement of these studies is provided by past studies, which have resequenced pharmacogenes in Africans and consistently discovered novel variation that is likely to affect treatment outcome [299–305]. As most African pharmacogenetic studies have been performed focussing on candidate genes [230], there is much that future pharmacogenomic studies in Africa stand to gain from the implementation of WGES.

3.5 Critical evaluation of the variation present in pharmacogenes

Before large and expensive projects are designed and initiated, proof of the value of these studies is required. The most cost-effective way of doing this is to review data that is already publically available. Therefore, as this article is focussed predominantly on the application of pharmacogenomic data, we critically examined and compared the variation present in the pharmacogenes of African and non-African individuals. In order to obtain a comprehensive overview of this variation, we utilised high throughput data derived from genome-wide studies utilising either large scale SNP genotyping or WGES.

To evaluate the value of WGES data, we utilised the Galaxy library and corresponding tutorial [306] to compare the variation observed in the pharmacogenes of thirteen ethnically diverse individuals, which were broadly divided into seven Africans and six non-Africans. For more details regarding the analyses performed and the individuals utilised, refer to Appendix 1 and Table S1. More specifically, we examined the top ten pharmacogenes (*ABCB1*, *CYP1A2*, *CYP2B6*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP3A4*, *TPMT*,

UGT1A1 and *VKORC1*) as described by Thorn *et al.* [307]. Although these pharmacogenes have all been extensively studied, using the data obtained from the Galaxy library for the thirteen individuals, we found that half of these genes contained novel/uncharacterised non-synonymous SNPs that were predicted to be “damaging” by the SIFT algorithm [308] (Table 3.1). In the context of this article we defined uncharacterised variants as those variants, which although predicted to alter the respective protein products, have not been described in either the allele nomenclature websites [309,310] or on PharmGKB [21] (accessed 1 February 2011). The presence of these variants emphasises the variability of the genes and suggests that more comprehensive genotyping assays may be required in future studies. More specifically, not only was there consistently a higher level of variation present in the pharmacogenes of African individuals when compared to non-Africans, but approximately three quarters of the novel/uncharacterised variation was detected in African individuals (Table 3.1). These data provide a strong argument for the resequencing of pharmacogenes, specifically in African descent individuals. The data also provide a potential explanation for the higher frequency of genotype-phenotype discordance observed in individuals with African ancestry [311], which may be attributed to uncharacterised variation present in pharmacogenes.

Table 3.1: Novel/uncharacterised non-synonymous SNPs predicted to be damaging, from resequencing data of 10 pharmacogenes in 13 ethnically diverse individuals

Gene	Africans	Non-Africans
<i>CYP1A2</i>		T324I
<i>CYP2C9</i>	L287M	
<i>CYP2C19</i>	L17H	
<i>CYP2D6</i>		R365H
	I118F	
<i>CYP3A4</i>	Q200H	
	L401P	

It is, however, equally important to consider variation that has been characterised and proven to be of value to pharmacogenetic applications. Therefore, we examined variation known to affect drug response, in the highest ranked genes with perceived importance for gene-drug interactions, as described by the CPIC [168] (Table 3.2). We also included *COMT* and *TPMT*-cisplatin to this list of drug-gene interactions, as it has recently been reported to be highly significant for pharmacogenetic applications [312]. To evaluate and compare the presence and frequency of this variation in Africans [Yoruba individuals from Nigeria (YRI)] and non-Africans [Northern and Western European ancestry individuals from Utah (CEU)], we utilised data from the International HapMap Project [313]. Unfortunately, consultation of the HapMap data revealed that genotype information was absent for almost half of the variants listed in Table 3.2, emphasising a need for population based studies focussing specifically on pharmacogenetic variation.

Table 3.2: List of important previously characterised variants affecting treatment response

Gene	Drug	Variant	ADR/treatment outcome
<i>CYP2D6</i>	Tamoxifen	*3 (rs35742686) ¹	Unfavourable treatment outcome
		*4 (rs3892097) ¹	
		*5 (gene deletion) ¹	
		*6 (rs5030655) ¹	
		*7 (rs5030867) ¹	
<i>CYP2C19</i>	Clopidogrel	*2 (rs4244285)	Decrease in platelet responsiveness
		*3 (rs4986893) ¹	Decrease in platelet responsiveness
		*17 (rs12248560)	Improved response
<i>VKORC1</i>	Warfarin	-1639 G>A (rs9923231)	Increased risk of bleeding
<i>CYP2C9</i>		1173 C>T (rs9934438)	Increased risk of bleeding
<i>CYP2C9</i>		*2 (rs1799853)	Increased risk of bleeding
		*3 (rs1057910)	Increased risk of bleeding
<i>HLA-B</i>	Abacavir	*5701 (rs2395029)	Hypersensitivity syndrome
<i>TPMT</i>	Mercaptopurine	*2 (rs1800462) ¹	Toxicity
		*3 (rs1142345)	Toxicity
		*4 (rs1800584) ¹	Toxicity
<i>TPMT</i>	Cisplatin	rs12201199	Hearing loss
<i>COMT</i>		rs9332377	Hearing loss

1: No genotype information available on HapMap

Even with the limited data, clear differences in allele frequencies between the two HapMap populations were observed (Figure 3.2A). In general, the variants appeared to be more representative of the CEU population, highlighting the disproportionate research that has been performed with regard to the characterisation of pharmacogenetic variation. Although not all of the relevant pharmacogenetic variants were detected in the African descent individuals, these variants have for the most part been well characterised and are thus likely to be accounted for by already available pharmacogenetic genotyping assays. Of further interest, although many variants were not detected in African descent individuals, some variants were present at much higher frequencies in these individuals and may therefore have even more value for pharmacogenetic applications in Africa (e.g. *COMT* rs9332377 and *TPMT* rs12201199, both associated with cisplatin-induced hearing loss [312]). The differences in allele frequencies observed between the YRI and CEU can be extended to the other nine HapMap populations, as has been well demonstrated by Adeyemo and Rotimi [58], who showed that the more removed the HapMap populations were from one another, the more the allele frequencies differed. Together these results serve to support the notion that multiple populations need to be studied in order for us to gain a comprehensive understanding of human variation and its effect on pharmacogenetic traits.

More specifically, even those studies that have taken African variation into account have predominantly used the YRI as a reference for African individuals. This is not always an accurate representation, as demonstrated by Rotimi and Jorde [314] with regard to the differences in allele frequencies observed for the *HLA-B**5701 abacavir hypersensitivity syndrome associated variant, and further highlighted in Figure 3.2B for *VKORC1* -1639 G>A, which is included in three of the five FDA approved *in vitro* diagnostic tests for warfarin

dosing [305]. Both these variants show that although the allele frequencies were extremely low in the YRI population, the same is not true for other African populations. Therefore, if a particular variant is not of relevance to one African population, it should not be assumed that it will not be relevant for pharmacogenetic applications in another African population. This once again highlights that the handful of African populations that have been included in GWAS and resequencing projects are unlikely to be representative of the entire continent.

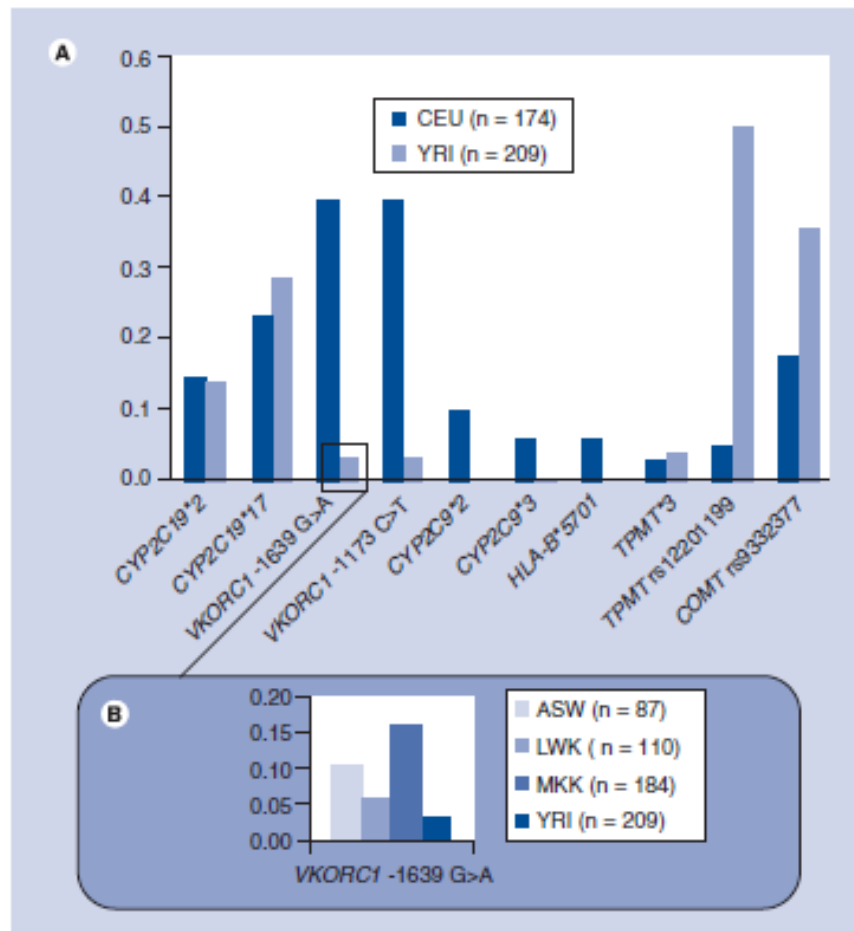


Figure 3.2 Differences in pharmacogenetic allele frequencies **A**: The differences in allele frequencies observed between the CEU and YRI individuals for known pharmacogene variants that have been genotyped by the HapMap Project. **B**: The differences in allele frequencies observed between African ancestry individuals for *VKORC1* -1639 G>A. Population descriptors: CEU: Utah residents with Northern and Western European ancestry from the CEPH collection, ASW: African ancestry in Southwest USA, LWK: Luhya in Webuye, Kenya, MKK: Maasai in Kinyawa, Kenya, YRI: Yoruban in Ibadan, Nigeria. All data were obtained from the HapMap Genome Browser release #28 [313].

In summary, these data show that available pharmacogenetic information has greater relevance for non-African populations, which may be attributed to the greater amount of research that has been performed, in combination with the lower levels of genetic variation and greater homogeneity of these populations, when compared to Africans. In contrast, not only is there a large proportion of uncharacterised variation present in African genomes, but the value of characterised variation differs between populations. These data all support the resequencing of many, diverse African genomes. Although we have demonstrated that

resequencing applications have the greatest value for African populations, as sequencing becomes cheaper and more readily available, the replacement of other less comprehensive genotyping assays with this technology would be beneficial to all population groups.

3.6 Hurdles associated with sequencing technologies

Before large scale resequencing projects are initiated, it is important that we are realistic about the potential hurdles associated with these new technologies. Although the time and costs associated with the generation of sequence data are steadily decreasing, the production of vast amounts of data offers a whole new range of constraints. A few years ago, the generation of sequence data was considered a rate-limiting step, now the analyses of these data create a variety of bottlenecks of their own. With regard to the cost of sequencing, it is often cheaper to resequence the data than to store it [315]. Thus, it is crucial that the technologies and techniques to store and analyse the generated data are developed at a rate equal to, or faster than, the rate at which sequencing technologies are developing. These additional costs and considerations may prevent the application of resequencing technologies in certain contexts and alternative genotyping strategies may need to be considered based on the resources available to the specific research unit. These genotyping assays may include commercially available pharmacogenetic tests, such as the AmpliChip® CYP450 Test [157] from Roche or the DMET™ chip from Affymetrix [316]. It should, however, be noted that these assays have predominantly been designed according to populations of European descent and may therefore not account for African specific alleles [299,303]. Although customised assays that are designed according to African populations will address these issues, they will most likely not account for rare variants, which are more common in African individuals [299,303]. Sequencing of candidate genes will provide data for all variants and will eliminate the bottlenecks caused by the vast amounts of data generated by WGES, however, this strategy will also eliminate the unbiased results that can be obtained by examining the entire genome. Therefore, it is essential to consider both the research question and the resources available before implementing a genotyping assay.

On a technical level, NGS technologies are far from perfect. However, the ability to detect variation has already improved dramatically since NGS was first introduced. Utilising the resequencing data for the thirteen individuals described in Table S1, we divided the individuals into eight individuals sequenced prior to 2010 and five individuals sequenced in 2010. Examination of the three *CYP2C19* alleles described in Table 3.2 showed that the ability to detect these variants differed according to the date that the sequencing data was published (Table S1 and Figure 3.3). This clearly demonstrates the improvement, not only in cost but also in quality, of NGS technologies (including techniques for the targeted capture of exomes) in a short period of time, the advancement of which is likely to continue. As technologies continue to improve in accuracy and decrease in cost, the likelihood of WGES becoming the genotyping assay of choice in the future will increase.

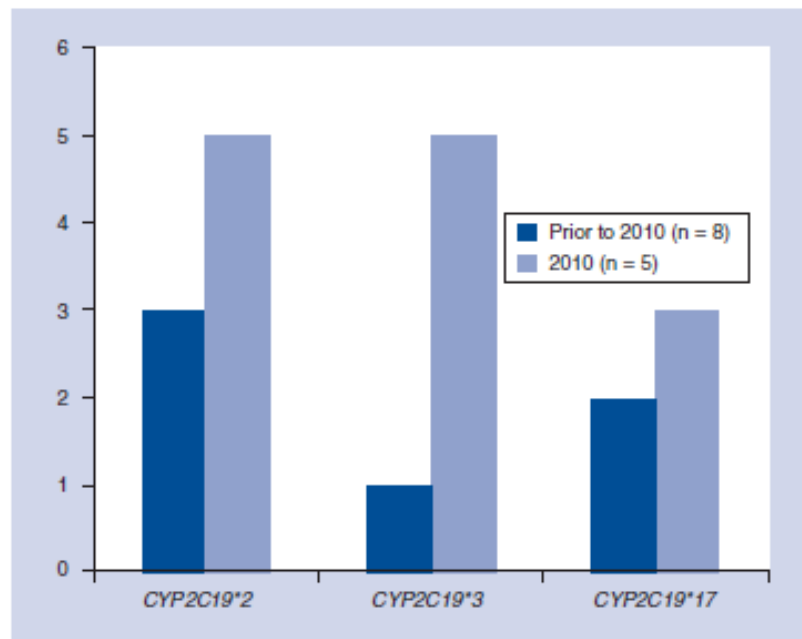


Figure 3.3: A comparison of the number of individuals for which genotype information was not available for the examined SNPs, stratified according to the date that the sequencing data was published. For the sequencing data published prior to 2010, eight individuals were examined, while for the sequencing data published in 2010, five individuals were examined. *CYP2C19*17* occurs beyond the regions captured for exome sequencing, thus explaining the inability to detect this variant in the two individuals for whom only exome sequences were published in 2010 and no whole genome data was available.

However there is still room for improvement, which is demonstrated by examining genotype data for *CYP2D6*, which is arguably one of the most important pharmacogenes [168]. The complex nature of the gene has resulted in both SNP-based and NGS technologies having difficulty capturing *CYP2D6* gene variation. This may in part be attributed to the high levels of sequence similarity observed between *CYP* genes and their corresponding pseudogenes [317] and is demonstrated by the lack of data pertaining to *CYP2D6* in the examined databases. In the HapMap database [313] the allele frequencies for only two SNPs are available and neither of these SNPs are of relevance to pharmacogenetic applications. Although genotype data for more SNPs are available in the VIP and 1000 Genomes Project database [318], this genotype data is far less than that obtained for other less variable genes (e.g. there are 82 *CYP2D6* alleles reported to date, which is more than double the number of alleles reported for the second most variable *CYP* gene present in the VIP genes and 1000 Genomes Project database).

Furthermore, the complexity of *CYP2D6* makes accurate allele prediction difficult. To demonstrate this, we reviewed the *CYP2D6* alleles reported for Craig Venter and James Watson by Ng *et al.* [319], to evaluate the potential for error in allele classification. We additionally included *CYP1A2*, *CYP2C9*, *CYP2C19* and *CYP3A4* in this evaluation of allele classification, as these genes fell into our previous criteria for the top ten pharmacogenes. Out of the ten predicted allele combinations for these five genes, we disagreed with four of the allele combinations reported by Ng and colleagues [319] (Table 3.3). For *CYP2C9* the discordance is of little consequence as both allele combinations result in EM phenotypes.

With regard to *CYP2C19*, however, we predicted that Craig Venter would be a *CYP2C19* PM (confirmed by the NCBI dbSNP database), while Ng and colleagues [319] predicted that he would be an EM. Additionally when looking at James Watson's genome, the presence of *CYP2C19*27*, which was reported on the *CYP* allele nomenclature website [309] post the Ng *et al.* [319] publication, serves as a reminder that new data are continually being discovered and should be added to databases on a regular basis. The *CYP2D6* data is, however, the most convincing with regard to the ease with which errors can occur in the allele classification process. Ng and colleagues [319] predicted that James Watson was homozygous for *CYP2D6*10*, which correlates with the resequencing data, such that he was homozygous for the P34S and S486T SNPs, which form the decreased function *CYP2D6*10* allele. However, when examining the data more extensively, we also noticed that there was no genotype information available for the 1846 G>A splicing defect variant, which along with the P34S and S486T SNPs, form the *CYP2D6*4* null allele - the most frequent non-functional allele in Caucasians [311]. When we consulted NCBI's Entrez SNP database [320], we found that James Watson was indeed homozygous for the 1846 G>A splicing defect and was thus a PM rather than an IM. Another consideration with regard to *CYP2D6*, is that any individual genotyped as homozygous for an allele, as both Craig Venter and James Watson were, may instead be hemizygous for that allele due to the presence of the *CYP2D6*5* gene deletion allele. Thus, for the time being, it appears that accurate *CYP2D6* allele prediction will require specially designed genotyping assays which will cover the spectrum of variation of this gene, accounting for the gene's homologous nature and including gene deletions and duplications, as well as hybrid genes. Furthermore, with regard to the assignment of pharmacogene alleles it appears that allele calling software, to eliminate potential human error, may be required. It is, however, important to remember that human interpretation is essential in order to critically examine the data and interpret the findings.

Table 3.3: Allele combination comparisons utilising data from our analyses to compare to that of Ng *et al.* [319].

Gene	Craig Venter		James Watson	
	Ng <i>et al.</i> [319] alleles	Our alleles	Ng <i>et al.</i> [319] alleles	Our alleles
<i>CYP1A2</i>	*1F/*1F	*1F/*1F	*1F/*1F	*1F/*1F
<i>CYP2C9</i>	*1A/*1B	*1A/*1B	*1A/*1A	*1A/*1C
<i>CYP2C19</i>	*1B/*1B	*2/*2	*1B/*1B	*1B/*27 ¹
<i>CYP2D6</i>	*1A/*1A	*1A/*1A	*10/*10	*4/*4
<i>CYP3A4</i>	*1A/*1A	*1A/*1A	*1A/*1B ¹	*1A/*1B ¹

Alleles in grey blocks are the combinations which do not correlate.
 1: The genotype given by the NCBI SNP database does not correlate to the given genotype, highlighting inconsistencies between different databases.

3.7 Conclusions

Genomic technologies have developed at a rapid pace. Unfortunately, although the technologies provide opportunities to aid in the improvement of human health, not all populations have benefitted equally from these technologies. The vast majority of genomic studies have been performed in European descent populations, followed by Asian populations. Even though the genetically diverse and ancient African populations are most

representative of all populations, they have been consistently under-represented in genomic research. Although it is generally accepted that Africa should be included in future studies, even with the advent of NGS technologies which are uniquely suited to the complex genomes of Africans, Africa remains neglected. NGS provides us with the much needed opportunity to catalogue human variation on a scale that would have been unimaginable a few years ago. Therefore we should take advantage of these opportunities and include all human populations in global genomic research endeavours.

It is, however, essential that we remain realistic with regard to what can be achieved both with reference to the technical challenges associated with current sequencing technologies and the translation of results. Sequencing technologies are still developing and there remains much room for improvement. This paper has shown that resequencing technologies are not yet capable of detecting all variation in all gene regions. The complex nature of certain genes, such as *CYP2D6*, are problematic for present day technologies, which have not yet completely overcome obstacles such as high sequence similarity. Additionally, the almost overwhelming amounts of data allow ample opportunity for the incorporation of errors, both human and technical. Therefore, it is important that we develop methods to analyse these data in an efficient and reliable manner that can be applied to the clinical setting.

Furthermore, if sequencing projects are to be implemented within Africa, with its lack of resources and infrastructure, there are several improvements that need to be made. These improvements include the development of appropriate facilities, collaborations and educational programs, all of which require a large amount of financial support. It is for this reason that the Human, Hereditary and Health in Africa (H3Africa) initiative has been proposed. This initiative, focussing on genomic research in Africa, aims to build resources and infrastructure within the continent, and in so doing contribute to the education and training of African scientists [293]. However, additional funding from organisations such as the government and private sectors is required. In order to convince these organisations of the value of the proposed research, proof of the clinical utility of genomic studies should be provided. Although, it may be argued that the translation of results into the clinical context has been slow, there have been a few good examples with regard to pharmacogenomics (as described in Chapter 1), which is highlighted by the fact that this field has been included as one of the suggested research avenues for the H3Africa initiative [293]. Thus, pharmacogenomics may be a good starting point for the utilisation of WGES for implementation in the clinic.

The question remains whether, at present, personalised medicine can be applied to every individual, all around the world. The short answer to this is, 'No'. At present, this is not feasible, especially in developing countries. Firstly there is not enough evidence to convince medical practitioners of the value of personalised medicine. Secondly, there is not enough information pertaining to the variation present in poorly characterised populations to confidently prioritise variants in pharmacogenes for genetic tests, the costs of which remain too expensive for routine use. Furthermore, sequencing technologies still need to develop, both with regard to quality and cost, to the point where it can be used in the clinical context. Sequencing technology is, however, in its infancy and if we look at the variation that has already been detected in the genomes of only five resequenced southern African

individuals [280], it is clear that there remains much to be discovered from the analyses of African genomes.

3.8 Future perspectives

In the future, WGES is likely to be applied more regularly in research. In five to ten years some of the main hurdles associated with present day sequencing will have been overcome. As the technology improves, the quality will improve and with this the reliability of variant detection. With regard to analysis of the data, optimised pipelines to suit the needs of the specific research project will be available, and computational training will be incorporated into biological degrees. Institutes making use of multi-disciplinary collaboration will become commonplace and large research networks, allowing for the creation of large cohorts, will be the norm. Furthermore, as the quality and reliability of NGS technologies improve and the costs decrease, these assays may replace the traditional SNP-based genotyping assays and will be of more value for detecting uncharacterised and rare variation, which has particular relevance for the highly diverse genomes of African individuals. However, although sequencing technologies may become more commonplace, it is important to remember that their use in research will depend largely on the context to which they are applied. Even though the associated sequencing costs are likely to continue to decrease, the costs and resources required to store the vast amounts of data may still limit the application of these technologies. Therefore before applying sequencing technologies, it will be necessary to carefully consider whether the large amounts of data that will be obtained are necessary, or reliable enough, to answer the research question of interest.

With regard to the sequencing data that is generated in the future, there are three main areas that we feel will be positively affected: (i) databases, (ii) the understanding of the genome and complex disorders, and (iii) incorporation of genomic data into the clinical setting. An increase in data will result in an increase in what is known regarding variation and as this data is added to the growing databases, it will become easier to correlate this variation to a specific phenotype or disease. More comprehensive databases will allow for more comprehensive studies and with this our knowledge of the genome will improve. This increase in understanding with regard to the genome and the complex interactions that exist within this system, in combination with the ability to detect all variation, including rare and novel variation, will give us insight into complex disorders. Lastly, as WGES becomes more reliable and easier to implement, we will reach a point where the whole genome sequence of an individual will become part of the medically relevant information of a patient and will be used to guide, among other things, treatment regimes.

Financial disclosure/acknowledgements:

The authors would like to acknowledge the South African NRF and Stellenbosch University for financial assistance. The authors have no other relevant affiliations or financial involvement with any organisation or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. No writing assistance was utilised in the production of this manuscript.

Declaration by the candidate:

With regard to Chapter 3 (p 40-52), the nature and scope of my contribution were as follows:

Nature of contribution	Extent of contribution (%)
Design of project Mining of African GWAS data Prioritisation of genes and variants of importance to the field of pharmacogenomics Identification and annotation of the variation present in the prioritised pharmacogenes in the “Bushman” library genomes Allele frequency comparisons of prioritised variants using HapMap data Examination of the improvement in the quality of whole genome sequence data over time Examination of <i>CYP2D6</i> variation data Critical assessment of the <i>CYP</i> alleles previously predicted for James Watson and Craig Venter Writing of article	85%

The following co-authors have contributed to Chapter 3 (p 40-52):

Name	Email address	Nature of contribution	Extent of contribution (%)
Prof Louise Warnich	lw@sun.ac.za	Design of project Critical assessment of article	10%
Dr Galen Wright	galen@sanbi.ac.za	Design of project <i>CYP</i> allele calling assessment Critical assessment of article	
Prof Robin Emsley	rae@sun.ac.za	Critical assessment of article	5%
Prof Dana Niehaus	djhn@sun.ac.za		

CHAPTER 4: Next generation sequencing of pharmacogenes: a critical analysis focusing on schizophrenia treatment

Drögemöller Britt I¹, Wright Galen EB^{1,2}, Niehaus Dana JH³, Emsley Robin³, Warnich Louise¹

¹ Department of Genetics, Stellenbosch University, South Africa

² South African National Bioinformatics Institute, University of the Western Cape, South Africa

³ Department of Psychiatry, Stikland Hospital, Stellenbosch University, South Africa

Reprinted with permission from Wolters Kluwer Health

4.1 Summary

Introduction: Due to the unmet needs of current pharmacotherapy for schizophrenia, antipsychotic pharmacogenetic research is of utmost importance. However, to date, few clinically applicable antipsychotic pharmacogenomic alleles have been identified. Nonetheless, next generation sequencing technologies are expected to aid in the identification of clinically significant variants for this complex phenotype. The aim of this study was therefore to critically examine the ability of next generation sequencing technologies to reliably detect variation present in pharmacogenes. **Methods:** Candidate antipsychotic pharmacogenes and very important pharmacogenes were identified from the literature and the Pharmacogenomics Knowledgebase. Thereafter, the percentage sequence similarity observed between these genes and their corresponding pseudogenes and paralogues, as well as the percentage low complexity sequence and GC content of each gene, was calculated. These sequence attributes were subsequently compared to the “inaccessible” regions of these genes as described by the 1000 Genome Project. **Results:** It was found that the percentage “inaccessible genome” correlated well with GC content ($P=9.96 \times 10^{-5}$), low complexity sequence ($P=0.0002$) and the presence of pseudogenes/paralogues ($P=8.02 \times 10^{-7}$). In addition it was revealed that many of the pharmacogenes were not ideally suited to next generation sequencing due to these genomic complexities. These included the *CYP* and *HLA* genes, both of which are of importance to many fields of pharmacogenetics. **Conclusions:** Current short read sequencing technologies are unable to comprehensively capture the variation in all pharmacogenes. Therefore, until high-throughput sequencing technologies advance further, it may be necessary to combine next generation sequencing with other genotyping strategies.

Key Words:

Antipsychotics, genome sequencing, pharmacogenomics, pharmacogenes, schizophrenia

4.2 Introduction

As detailed in Chapters 2.1 and 2.2 schizophrenia is one of the most debilitating mental disorders and current antipsychotic treatments have substantial limitations with poor response rates [186], high relapse rates [101] and many severe forms of ADRs [119]. Antipsychotic treatment response varies substantially between individuals [218], highlighting the need for accurate genetic predictors of this phenotype. Furthermore, both schizophrenia and antipsychotic treatment response have been shown to be highly heritable [46,171] and for this reason it seems likely that genetic variation plays an important role in antipsychotic treatment outcomes.

Although there was much initial anticipation regarding the application of antipsychotic pharmacogenetics [171], it appears that both schizophrenia and antipsychotic treatment response are complex phenotypes. Thus, it seems more likely that rare variants and/or several common variants in many genes will interact with one another to influence the range of phenotypes that are observed with regards to both the disorder and treatment thereof [321]. This genomic complexity, combined with environmental influences, may explain the lack of clinically useful results that have been obtained from antipsychotic pharmacogenetic studies to date. Adding to this void, most of the past studies have focused on examining variants in single candidate genes (refer to Chapter 2.3.2 for more details) and have been largely unable to simultaneously examine the variation present in all the genes in the genome (refer to Table S2 for antipsychotic pharmacogenetic review articles). By analysing the entire spectrum of variation present in entire gene networks and pathways, we may be able to obtain a more comprehensive overview of the genetic factors contributing to antipsychotic treatment phenotypes. This may be achieved through the implementation of NGS technologies, which allow for the high throughput analyses of all variants in all genes.

Unfortunately, although these sequencing technologies have revolutionised the field of genomics in a remarkably short time, the human genome remains complex containing regions which are repetitive, GC rich and/or exhibit areas of high sequence similarity. These complexities hinder NGS technologies by decreasing the accessibility of the genome or interfering with the alignment of sequence reads [322–324]. Currently the most affordable NGS technologies for whole genome analyses still utilise short read sequencing. Consequently areas of high sequence similarity are often affected by misalignments, even though there have been large advances in alignment and variant calling algorithms such as those implemented by the Burrows-Wheeler Alignment (BWA) Tool [325] and the Genome Analysis Toolkit (GATK) [326]. The inability of these sequencing technologies to capture all the variation in the entire genome, without bias, is reflected by differences in sequencing coverage across the genome.

The 1000 Genomes Project, which has made excellent use of NGS to characterise the human variome across different world population groups, has drawn attention to the fact that differences in sequence coverage across the genome may act as an indication of which areas of the genome are accessible to short read sequencing [281]. Variants that occur outside of these accessible areas may not always be reliably called. In the pilot phase of the 1000 Genomes Project, the “accessible genome” was calculated by determining which areas contained coverage that differed by a factor of 2 from the median coverage across the genome, as well as which areas had more than 10% of their reads exhibiting mapping quality scores of less than 0 [327]. All areas falling outside of these specifications were considered accessible [281]. After the completion of the sequencing of 1 092 genomes from 14 populations, the 1000 Genomes Project also created a more stringent definition of the “accessible genome”. This stated that the coverage of an area needed to be within 50% of the average coverage across the genome, only 0.1% of the reads could have mapping quality scores of 0 and the average mapping quality needed to be greater than 56 [71]. Although this definition is stringent and only 2% of sites that are called using the GATK’s variant quality score re-calibration are likely to be false positives [71], using this strict mask may

serve as a warning for problematic areas and variants called in these areas may need to be examined with caution.

The occurrence of inaccessible areas within the genome may be of particular relevance to the field of pharmacogenomics. Some of the most important pharmacogenes include the *CYP* and *HLA* gene families, both of which have been implicated in antipsychotic pharmacogenetics (Table S2). These genes are highly polymorphic, with the *HLA* region being the most polymorphic in the human genome [328], and variation within these genes are documented and analysed using special nomenclature systems [309,329,330]. For this reason the variation present in these families of genes requires extensive characterisation, highlighting the utility of sequencing technologies as genotyping strategies for these genes. Unfortunately, these genes also show areas of high sequence similarity to other regions of the genome due to the large number of related genes and pseudogenes [330,331]. This draws attention to the likelihood that the *CYP* and *HLA* genes may not be well suited to NGS (the difficulties associated with *CYP2D6* have already been alluded to in the previous chapter) and it remains likely that other pharmacogenes may be affected in a similar manner.

The analyses performed by this study aimed to assess the ability of short read NGS technologies to reliably detect the variation present in pharmacogenes related to the antipsychotic treatment of schizophrenia. Furthermore, we examined pharmacogenes, which have been shown to be most relevant to the field of pharmacogenetics in the broader sense, and compared them to the antipsychotic pharmacogenes. The assessment of all the pharmacogenes was performed by critically examining sequence coverage data in combination with the genomic complexities present in these gene regions.

4.3 Materials and methods

4.3.1 Identification of candidate antipsychotic pharmacogenes and very important pharmacogenes

In order to identify candidate genes that are of interest to schizophrenia related antipsychotic pharmacogenetics, a literature search was performed in the PubMed Database using the search terms “antipsychotic pharmacogenetics” and “antipsychotic pharmacogenomics” [82] (accessed 15 January 2013). To ensure that the latest and most relevant candidate genes were identified, “review” was used as an article type filter and a publication date filter of “5 years” was incorporated into the search. References of identified articles were reviewed for additional relevant citations. Articles that were not available in English and were not related to genetic association studies examining the antipsychotic treatment of schizophrenia were excluded. The remaining articles were subsequently mined to identify genes that are annotated on the reference sequence and have been associated with antipsychotic treatment response or ADR phenotypes of relevance to the treatment of schizophrenia.

In addition to the identified antipsychotic pharmacogenes, all PharmGKB VIPs [170] (accessed 15 January 2013) were included in downstream analyses. These genes were included to serve as a comparison for the antipsychotic pharmacogenes. Furthermore, the inclusion of pharmacogenes of high relevance to other fields of pharmacogenetics allowed

for the results obtained from this study to have relevance to the field of pharmacogenetics as a whole.

4.3.2 Critical analyses of factors potentially influencing the sequencing of pharmacogenes

Three main factors that could potentially impact results obtained from the NGS of the identified candidate genes were considered and the detailed methodology for the analyses used to examine these factors can be found in Appendix 2. These three factors were (i) high sequence similarity to paralogues or pseudogenes, (ii) GC content and (iii) repetitive or low complexity sequences. High sequence similarity may result in misalignment of the sequence reads and GC content may affect the accessibility of gene regions for sequencing applications, while low complexity sequences may affect both these aspects.

Identification of paralogues and pseudogenes

Paralogues showing greater than 70% sequence similarity to the candidate genes of interest were identified through the use of Ensembl BioMart [332] using the Ensembl Genes 69 Database and the *Homo sapiens* genes (GRCh37.p8) dataset. Related pseudogenes, or related functional genes in cases where the candidate genes were pseudogenes, were identified and gene sequences were obtained using NCBI's gene resource [333]. To determine which of the pseudogenes contained areas with more than 70% sequence similarity to the genes of interest, mVISTA was used [334].

Calculation of percentage GC content and low complexity sequences

The GC content of each gene was calculated through the utilisation of Ensembl BioMart [332]. To identify the percentage of low complexity or repetitive sequence present in each gene, the gene co-ordinates were obtained from Ensembl BioMart [332]. These co-ordinates were then used to determine the percentage of masked sequence present in each gene by using the Pre-masked Genome Search, available from RepeatMasker [335].

4.3.3 Assessment of pharmacogenes using the 1000 Genomes Project mask files

Once candidate antipsychotic pharmacogenes and VIPs had been identified and critically examined, the "accessible genome", as defined by the 1000 Genomes Project coverage data [71], was used as a proxy for the ability of these genes to be successfully sequenced. In order to determine which areas of the candidate genes did not fall into the "accessible genome" (referred to as the "inaccessible genome") as calculated by the 1000 Genomes Project "strict mask", the bed file containing the unmasked areas was downloaded [327]. Thereafter, the areas falling into the gene regions as defined by Ensembl BioMart [332] were assessed to determine how many base pairs fell outside of the "accessible genome". Refer to Appendix 2 for more details.

4.3.4 Statistical analyses

Differences between the VIPs and antipsychotic pharmacogenes with regards to percentage inaccessibility, GC content and low complexity sequences were assessed employing the Wilcoxon-Matt-Whitney test. A Pearson's Chi-squared test of independence was used to assess the differences between these two groups with regards to the number of genes with more than 70% sequence similarity to pseudogenes/paralogues. The relationship between percentage inaccessibility and the dichotomous presence of a pseudogene or paralogue

(with more than 70% sequence similarity) was also assessed using the Wilcoxon-Matt-Whitney test. Additionally, associations between percentage inaccessibility, GC content and low complexity sequences were examined using the Spearman's rank-order correlation. For group comparisons, genes belonging to both the VIP and antipsychotic pharmacogene lists were excluded from analysis. *P* values less than 0.05 were considered significant. All statistical analyses were performed in R [336].

4.4 Results

Using the search criteria for antipsychotic pharmacogenetics literature as described above, 38 articles remained which referred to associations that were found with variants in 152 genes (Table S2). The antipsychotic pharmacogenetic traits that these genes were associated with were: treatment response, weight gain, movement disorders, agranulocytosis, QT prolongation, hyperprolactinemia and neuroleptic malignant syndrome. When the search was broadened to include important pharmacogenes from other fields of pharmacogenetics, it was found that PharmGKB listed 47 VIPs, of which 12 were also included in the list of antipsychotic pharmacogenes (Table 4.1).

Statistical analyses revealed significant positive correlations between percentage inaccessibility and GC content ($\rho=0.281$, $P=9.96 \times 10^{-5}$), as well as percentage low complexity sequence ($\rho=0.269$, $P=0.0002$). The mean percentage inaccessibility was higher for genes with paralogues/pseudogenes (i.e. 43.47%) compared to those without these homologous regions (i.e. 21.30%) and this difference was statistically significant ($P= 8.02 \times 10^{-7}$). The results obtained from the analyses of the genomic composition (GC content, sequence similarity, low complexity sequences and “inaccessible genome”) of the candidate antipsychotic pharmacogenes and VIPs are shown in Table S2. When examining the pharmacogenes that had more than 50% of their gene regions falling into the “inaccessible genome” (Table 4.2), 23 genes were identified. Examination of the genomic composition of these genes revealed that 20 of the genes (86.96%) exhibited more than 70% sequence similarity to paralogues or pseudogenes. When examining the remaining three genes, it was observed that two of these three genes were affected by low complexity sequence or high GC content. In the case of the first gene, *TGFB1*, approximately half of the sequence present in this gene was repetitive (i.e. constitutes low complexity sequence). In the case of the second gene, *DRD4*, Figure 4.1 depicts how all but two of the inaccessible areas in *DRD4* are either repetitive or GC rich.

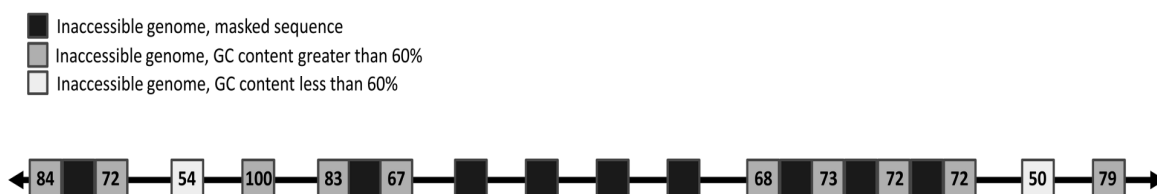


Figure 4.1: *DRD4* and the corresponding inaccessible regions of the gene. The regions are not drawn to scale and are merely a representation of how genomic complexities can influence genome sequencing coverage. The blocks represent areas of the gene that are inaccessible and are: (i) masked by RepeatMasker; (ii) contain greater than 60% GC content; and (iii) contain less than 60% GC content and are not masked by RepeatMasker. The numbers in the blocks indicate the GC content of these areas.

Table 4.1: The genomic composition of antipsychotic pharmacogenes that are also considered to be very important pharmacogenes

Gene	Associated Antipsychotic Pharmacogenetic Trait	Reference	Paralogue/Pseudogene	GC Content (%)	Low Complexity Sequence (%)	“Inaccessible Genome” (%)
<i>ABCB1</i>	Treatment response, Weight gain	[16,120,127,171,176,177,179–182]	> 70% sequence similarity	37.05	51.55	19.39
<i>ADRB2</i>	Weight gain	[178]		50.61	0.00	13.97
<i>COMT</i>	Treatment response, Movement disorder	[16,73,100,119–121,171,176,177,179,183,185,187,189,193,195–198]		53.41	50.18	39.19
<i>CYP1A2</i>	Treatment response, Movement disorder	[15,16,119,120,171,176,177,179,183–186]	> 70% sequence similarity	52.03	32.79	24.48
<i>CYP2C19</i>	Treatment response	[88,179]	> 70% sequence similarity	38.88	80.16	62.63
<i>CYP2D6</i>	Treatment response, Movement disorder, Weight gain, QT prolongation	[16,88,119–121,127,171,176,177,179–181,183–192]	> 70% sequence similarity	62.68	0.00	100.00
<i>CYP3A4</i>	Treatment response	[120,176]	> 70% sequence similarity	39.62	39.33	46.57
<i>CYP3A5</i>	Treatment response	[120]	> 70% sequence similarity	40.47	48.48	31.64
<i>DRD2</i>	Treatment response, Movement disorder, Weight gain, Hyperprolactinemia, Neuroleptic malignant syndrome	[16,73,88,100,119–121,127,148,171,176–181,183–187,189–191,193–196]		48.36	33.70	9.33
<i>GSTP1</i>	Movement disorder	[16,183]	> 70% sequence similarity	63.05	2.84	25.74
<i>MTHFR</i>	Treatment response, Weight gain	[16,176,177,179–181,195]		54.53	27.37	22.32
<i>NQO1</i>	Movement disorder, Agranulocytosis	[16,136,171,176,183,185]		47.09	54.04	47.77

Table 4.2: The genomic composition of those genes with more than 50% “inaccessible genome”

Antipsychotic Pharmacogenes						
Gene	Associated Antipsychotic Pharmacogenetic Trait	Reference	Paralogue/Pseudogene	GC Content (%)	Low Complexity Sequence (%)	“Inaccessible Genome” (%)
<i>DRD4</i>	Treatment response, Movement disorder, Weight gain	[16,73,119,171,176,177,179–181,183,185,190,191,193,195]		67.02	21.56	65.10
<i>GSTM1</i>	Movement disorder	[16,171,183–185]	> 70% sequence similarity	46.35	55.35	96.88
<i>HLA-B</i>	Agranulocytosis	[121,136]	> 70% sequence similarity	59.00	0.00	97.01

<i>HLA-C</i>	Agranulocytosis	[121,136]	> 70% sequence similarity	59.28	0.00	99.26
<i>HLA-DQA1</i>	Agranulocytosis	[121,136]	> 70% sequence similarity	39.77	44.87	77.50
<i>HLA-DQB1</i>	Agranulocytosis	[16,121,136,176,186,189]	> 70% sequence similarity	47.03	13.30	95.00
<i>HLA-DQB3*</i>	Agranulocytosis	[136]	> 70% sequence similarity	48.72	0.00	72.15
<i>HLA-DRB1</i>	Agranulocytosis	[121,136]	> 70% sequence similarity	40.74	33.79	99.84
<i>HLA-DRB5</i>	Agranulocytosis	[121,136]	> 70% sequence similarity	42.84	36.39	100.00
<i>HSPA1A</i>	Agranulocytosis	[121,136]	> 70% sequence similarity	59.76	0.00	97.37
<i>OXT</i>	Treatment response	[176]	> 70% sequence similarity	71.94	0.00	88.18
<i>TGFB1</i>	Weight gain	[181,184]		52.41	50.01	61.39

Antipsychotic Pharmacogene falling into the Very Important Pharmacogene Category

Gene	Associated Antipsychotic Pharmacogenetic Trait	Reference	Paralogue/Pseudogene	GC Content (%)	Low Complexity Sequence (%)	"Inaccessible Genome" (%)
<i>CYP2C19</i>	Treatment response	[88,179]	> 70% sequence similarity	38.88	80.16	62.63
<i>CYP2D6</i>	Treatment response, Movement disorder, Weight gain, QT prolongation	[16,88,119–121,127,171,176,177,179–181,183–192]	> 70% sequence similarity	62.68	0.00	100.00

Other Very Important Pharmacogenes

Gene	Associated Antipsychotic Pharmacogenetic Trait	Reference	Paralogue/Pseudogene	GC Content (%)	Low Complexity Sequence (%)	"Inaccessible Genome" (%)
<i>ADRB1</i>	NA	NA		57.51	6.99	51.80
<i>BRCA1</i>	NA	NA	> 70% sequence similarity	44.09	57.07	58.64
<i>CYP2A6</i>	NA	NA	> 70% sequence similarity	53.42	20.61	98.57
<i>CYP2B6</i>	NA	NA	> 70% sequence similarity	44.63	60.12	73.13
<i>CYP2C9</i>	NA	NA	> 70% sequence similarity	37.78	71.96	56.82
<i>GSTT1</i>	NA	NA	> 70% sequence similarity	52.22	44.13	65.34
<i>SULT1A1</i>	NA	NA	> 70% sequence similarity	52.72	44.23	89.67
<i>TYMS</i>	NA	NA	> 70% sequence similarity	45.77	49.32	52.92
<i>VKORC1</i>	NA	NA	> 70% sequence similarity	55.01	41.11	53.17

*This gene is a pseudogene.

Analysis of differences between the VIPs and antipsychotic pharmacogenes with regards to percentage GC content was not statistically significant ($P=0.915$), however the differences between the two groups with regards to percentage inaccessibility was statistically significant ($P=0.035$). Finally, it was shown that the VIPs were more likely to be affected by areas of high sequence similarity ($P=0.029$), with 42.86% of the VIPs exhibiting more than 70% sequence similarity to paralogues or pseudogenes (compared to 24.29% of the antipsychotic pharmacogenes).

4.5 Discussion

This article has identified a significant number of studies that have reported associations with various antipsychotic pharmacogenetic traits. The majority of the genes examined in these studies were associated with treatment response (52.26%), followed by weight gain (29.68%), movement disorders (23.23%) and agranulocytosis (9.03%) (Table S2). The most likely reasons for this are: (i) treatment response, although difficult to measure, is arguably the most significant hurdle in the treatment of schizophrenia and only approximately half of patients respond to treatment [186]; (ii) weight gain and related metabolic disorders are the most prominent ADRs with atypical antipsychotics [337]; (iii) movement disorders are the most frequent ADRs observed with typical antipsychotics [337] and (iv) agranulocytosis is a severe, and in some cases lethal, ADR associated with antipsychotic treatment [338] (refer to Chapter 2.2 for more details). These traits were associated with several different genes and interestingly, of the 47 VIPs, 12 (25.53%) were also antipsychotic pharmacogenes. This highlights the importance of antipsychotic pharmacogenetics in the context of pharmacogenetics as a whole. Thus, NGS projects examining antipsychotic pharmacogenomics may be valuable and therefore it is necessary to critically examine the likelihood that antipsychotic pharmacogenes will be successfully sequenced.

Although the 1000 Genomes Project masking may be an overly stringent assessment of the quality of sequencing data, the findings from this study indicate that the areas of the genome that were considered to be inaccessible correlated well with all three genomic complexities that were examined (percentage GC content: $P=9.96 \times 10^{-5}$, percentage low complexity sequence: $P=0.0002$, presence of paralogues/pseudogenes with more than 70% sequence similarity: $P=8.02 \times 10^{-7}$). This aligns with the description of the 1000 Genomes Project masked regions, which are described as areas where reads are “ambiguously placed” or where there are “unexpectedly high or low numbers of aligned reads” [281]. These discrepancies in read alignment may stem from the initial capture and amplification processes during the library preparation prior to sequencing. However, it remains likely that a portion of these reads are successfully captured, amplified and sequenced, but alignment of these sequence reads is error prone [339]. Therefore, the library preparation and alignment algorithms associated with NGS technologies may require improvement and it is for this reason that sequencing companies are addressing these issues [324].

With reference to the specific genes that were considered inaccessible, *DRD4* may be especially important to consider for antipsychotic pharmacogenomic studies, as all current antipsychotics bind to dopamine receptors [96]. This gene is affected by both low complexity sequence and high GC content (Figure 4.1) and these attributes are likely to affect the success with which this gene can be sequenced using NGS technologies. Along with this gene a further 22 pharmacogenes contained more than 50% inaccessible

sequence, with 20 of these genes showing greater than 70% sequence similarity to paralogues or pseudogenes and one gene containing approximately 50% low complexity sequences. In total, 22/23 (95.65%) of the genes that were considered largely inaccessible (more than 50% inaccessible), also contained a significant percentage of sequence complexities (more than 50% of their gene regions were affected by sequence similarity, high GC content or low complexity sequences).

When comparing the antipsychotic pharmacogenes and VIPs, it was observed that these two groups were similar with respect to percentage GC content and low complexity sequence. Although there was a significant difference in the percentage inaccessible sequence ($P=0.035$), this was likely to be driven by the significant differences that were observed with respect to the number of genes with more than 70% sequence similarity to pseudogenes/paralogues ($P=0.029$). As antipsychotic pharmacogenes are often required for vital processes such as dopamine regulation, they are more likely to be under evolutionary constraint. In contrast, the largest gene family present in the VIPs, namely the *CYP* gene family (10 of the 47 VIPs are *CYP* genes), has evolved rapidly. Furthermore, it has been hypothesised that as a result of the current non-essential nature of these genes, they have accumulated many variants [331]. The resulting large polymorphic gene families make short read sequencing of these areas very challenging and highlight the potential short-comings with regards to NGS in the context of pharmacogenomic studies (Chapter 3). The hurdles associated with the NGS of polymorphic gene families such as the *CYP* and *HLA* genes, are reflected in the large areas within these genes that are deemed inaccessible by the 1000 Genomes Project strict masking annotation. Both of these families contain genes that are 100% masked. These genes, namely the *CYP2D6* and *HLA-DRB5* (Table 4.2), show large regions of sequence similarity to other genes or pseudogenes within their respective families. This is clearly demonstrated in Figure 4.2, where the high sequence similarity observed between *CYP2D6* and the two corresponding pseudogenes is displayed.

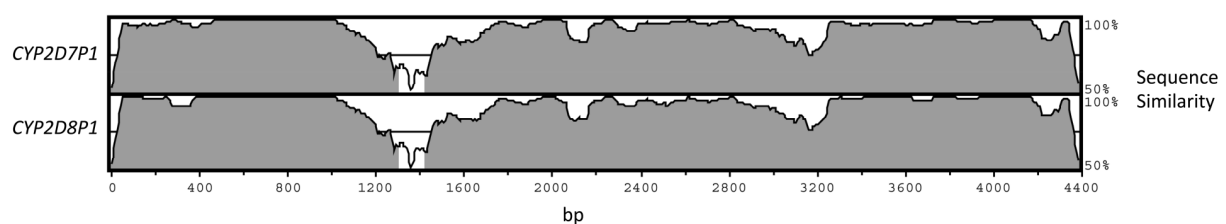


Figure 4.2: The percentage sequence similarity between *CYP2D6* and the corresponding pseudogenes as calculated by mVISTA.

These findings reiterate that the presence of sequence complexities may hinder the unbiased nature of WGES and it seems likely that certain areas within the genome may remain impervious to NGS variant detection, as was well demonstrated in a recent study examining kidney disease [323]. It is important to note that these sequence complexities not only complicate the identification of single nucleotide variants, but may also further hamper the detection of CNVs. This is of particular relevance to pharmacogenes such as *CYP2D6* where gene deletions, duplications and hybrid genes have been reported [173]. Although it is possible to examine CNVs with algorithms designed to examine NGS read depth, the genomic complexities examined in this study are likely to confound these results [340]. Thus, the examination of CNVs present in pharmacogenes that are affected by these sequence complexities is likely to yield inaccurate results.

It is well known that although NGS technologies far surpass Sanger sequencing with regards to time and cost [273], the accuracy of Sanger sequencing and CNV assays for variant detection remains unparalleled. From the results obtained in this study, it appears that current short read sequencing methods are not sufficiently reliable for variant detection for many pharmacogenes. This highlights the fact that the design of NGS pharmacogenetic assays, such as the PGRN-Seq [341], needs to be performed with caution. This assay warns that genes such as *CYP2D6*, *HLA-B* and *HLA-DQB3* are unlikely to be accurately sequenced due to high sequence similarities, which correlates well with our data. All three of these genes are antipsychotic pharmacogenes and *CYP2D6* is considered among the top ten VIPs [307], possibly limiting the applicability of this assay for antipsychotic pharmacogenetic applications. The difficulties with genotyping both *CYP2D6* and the *HLA* genes have been previously documented and carefully designed genotyping strategies are required, including CNV assays [304,328,329,342]. With particular reference to using NGS for *HLA* genotyping it appears that the use of RNA-seq, which allows for greater representation of *HLA* alleles that are highly expressed, in combination with longer read lengths, to prevent misalignment, may be a better strategy [328,329]. Even so variant detection in these genes will remain a challenge.

Although the genomic complexities associated with many of the pharmacogenes do provide unique challenges, the advancement of sequencing technologies offers the potential for the discovery of variants associated with antipsychotic response phenotypes. NGS approaches can be used to simultaneously examine known pathways, as well as to discover novel targets, while allowing for the detection of both common and rare variants. However, for NGS to be truly unbiased, it is essential that all genes are represented, including well known pharmacogenes such as the *CYP* and *HLA* families. The results from this study serve as a reference for which pharmacogenes may require careful analyses with regards to NGS data. It should, however, be noted that these results are limited as they only refer to candidate pharmacogenes and examine these genes as a whole. Future studies may need to examine additional genes and it may be necessary to focus on specific regions within these genes.

4.6 Conclusion

Future advances are likely to incorporate longer read lengths in sequencing technologies such as nanopore technology [343], thereby enabling comprehensive characterisation of variation in all genes. Studies making use of well characterised cohorts in combination with these comprehensive genotyping strategies may then obtain results that are applicable to clinical practice. In the interim, the combination of short read NGS analyses with already existing strategies, such as the use of long-range PCR or longer read NGS, may be the best strategy to examine antipsychotic pharmacogenetics.

Financial disclosure/acknowledgements:

BI Drögemöller: NRF research bursary and the L'Oréal-UNESCO for women in Science in Sub-Saharan Africa Fellowship. This financial assistance is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily attributed to these organisations.

Declaration by the candidate:

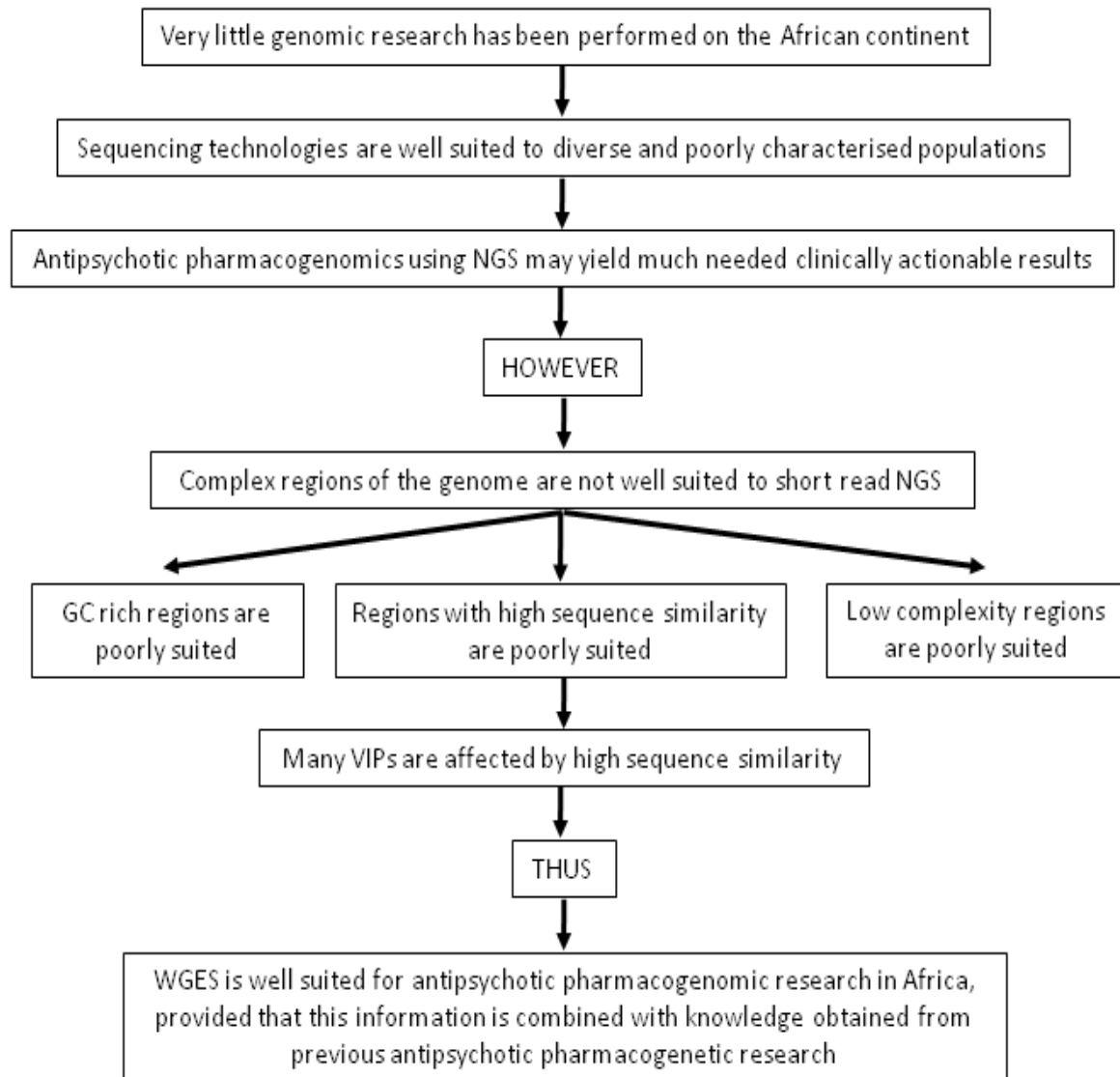
With regard to Chapter 4 (p 54-63), the nature and scope of my contribution were as follows:

Nature of contribution	Extent of contribution (%)
Design of project Literature search and mining of review articles Identification of candidate genes Identification of pseudogenes/paralogues Calculation of percentage sequence similarity Calculation of percentage GC content Calculation of percentage low complexity sequence Calculation of percentage "inaccessible" genome Critical analysis of genomic composition in relation to percentage "inaccessible" genome Writing of article	80%

The following co-authors have contributed to Chapter 4 (p 54-63):

Name	Email address	Nature of contribution	Extent of contribution (%)
Prof Louise Warnich	lw@sun.ac.za	Design of project Critical assessment of article	15%
Dr Galen Wright	galen@sanbi.ac.za	Design of project Calculation of percentage "inaccessible" genome Critical analysis of genomic composition in relation to percentage "inaccessible" genome Critical assessment of article	
Prof Robin Emsley	rae@sun.ac.za	Critical assessment of article	5%
Prof Dana Niehaus	djhn@sun.ac.za		

Summary of Chapters 3 and 4



PART 2: Utilisation of exome sequencing for antipsychotic pharmacogenomics in the South African context

CHAPTER 5: Patterns of variation influencing antipsychotic treatment outcomes in South African first episode schizophrenia patients

Drögemöller Britt I¹, Niehaus Dana JH², Chiliza Bonginkosi², Asmal Laila², Malhotra Anil K³, Wright Galen EB^{1,*}, Emsley Robin², Warnich Louise¹

¹Department of Genetics, Stellenbosch University, South Africa;

²Department of Psychiatry, Stikland Hospital, Stellenbosch University, South Africa;

³Department of Psychiatry, Zucker Hillside Hospital, United States;

*Current address: South African National Bioinformatics Institute, University of the Western Cape, South Africa

Reprinted with permission from Future Medicine.

5.1 Summary

Introduction: Many antipsychotic pharmacogenetics studies have been performed examining candidate genes or known variation, however, our understanding of the genetic factors in antipsychotic pharmacogenetic traits involved remain limited. **Materials and methods:** A cohort of 130 first episode schizophrenia patients was collected and treated with flupenthixol decanoate for twelve months, after which a subset of non-responders and responders to treatment were identified for exome sequencing. The variation observed in the responders and non-responders was subsequently compared. **Results:** Examination of coding variation revealed a potential role for rare loss of function variants in treatment response outcomes. In addition, the majority of the loss of function variation that was considered likely to be involved in non-optimal antipsychotic treatment response was either novel or rare in Asian and European populations. **Conclusion:** This study highlights the importance of exome sequencing for antipsychotic pharmacogenomics studies, particularly in African individuals, and the results suggest that future research should focus on rare loss of function variants in diverse populations.

Keywords:

Antipsychotic treatment, exome, genome, rare variants, sequencing, schizophrenia, South Africa

5.2 Introduction

As described in Chapter 3, there have been many advances in human genetics as a result of advances in sequencing technologies [25]. Nonetheless, there remain a large number of diseases for which substantial genomic and pharmacogenomic information is lacking, of which schizophrenia is one [54,344]. To aid in genomic research efforts to identify the variants involved in the schizophrenia, large consortiums such as the PGC [60] have been formed. Unfortunately, even though these consortiums are utilising thousands of patients and novel loci have been identified [61,62], there are still gaps in our knowledge regarding the genetics of schizophrenia (refer to Chapter 2.1.3 for more details regarding the current genetic findings).

Therefore, with the rapid decrease in sequencing costs [273], research efforts have turned to WGES for much needed answers. In contrast to the hundreds of candidate gene studies that have been performed [53], there are currently only six published schizophrenia studies using WGES strategies [65–70] and no WGES studies focusing on antipsychotic treatment response. As reported in Chapter 2.1.3, the results from the WGES schizophrenia studies have added to the growing evidence that there may be more variants, both *de novo* and rare, present in schizophrenia patients and that these variants are more likely to have an adverse effect on the resulting protein products. This corroborates earlier findings that schizophrenia patients are more likely to harbour damaging variants in the form of rare CNVs [345]. In addition to this, it has been reported that rare variants confer a bigger risk for schizophrenia susceptibility than common variants [52] (Figure 2.3). In contrast to GWAS, WGES is able to detect both common and rare variants, highlighting the importance of utilising such technologies when investigating schizophrenia and antipsychotic treatment response outcomes.

Although there are no large consortiums or WGES studies that have focused on antipsychotic pharmacogenomics, this avenue of schizophrenia research is important. As detailed in Chapter 2.2 antipsychotics are the only known therapeutic agents that are consistently shown to be superior to placebos for the treatment of schizophrenia [77,78], however, treatment outcomes are plagued by a lack of efficacy and many ADRs [84,85]. Reports that differences in treatment response are heritable [172] highlight the need to elucidate the genetic factors involved in influencing response to antipsychotic treatment. Unfortunately, past research has focused predominantly on candidate genes [176] and the majority of findings have not been well replicated [172]. Therefore, it seems likely that many genes and variants, some that could still be unknown, may be involved in antipsychotic treatment response outcomes, making WGES an ideal tool for antipsychotic pharmacogenomic research.

This study therefore sequenced the exomes of eleven South African FES patients to examine the patterns of variation contributing to antipsychotic treatment outcomes, focusing predominantly on functional variants. These results should play a role in enhancing our understanding of antipsychotic treatment response in the South African context and guide the design of future antipsychotic pharmacogenomic studies.

5.3 Materials and methods

5.3.1 Patient samples

Written informed consent was provided by all patients and/or their guardians prior to this study and ethical clearance was obtained from the Committee for Human Research, Faculty of Health Sciences, Stellenbosch University. A cohort of 130 South African FES patients, who were assessed with the Structured Clinical Interview for the DSM-IV [43], were collected over four years. After a wash out period of up to seven days, all patients were treated with flexible doses of flupenthixol decanoate and treatment response was measured by means of the PANSS for twelve months. Patients were monitored every two weeks for the first six weeks, and every three months thereafter.

From this cohort a subset of ten South African SAC patients falling at extreme ends of the treatment response phenotype were selected for exome sequencing. These patients

consisted of five non-responders and five responders to treatment. The non-responder individuals all discontinued treatment due to a lack of efficacy, while the responder individuals all exhibited a greater than 40% improvement in PANSS total scores during the twelve months of treatment. The non-responder and responder individuals were matched for age (within five years) and gender. In addition, to allow for the examination of shared variation between related individuals, a sibling of one of the non-responders was selected for exome sequencing. This sibling was also a non-responder to flupenthixol decanoate treatment.

5.3.2 Exome sequencing

Genomic DNA (gDNA) from the eleven individuals was extracted from venous blood using the QIAamp DNA Blood Maxi Kit (Qiagen, Germany), according to the manufacturer's instructions (refer to Appendix 3 for details). Thereafter, the gDNA samples were sent to the HudsonAlpha Genomics Service Laboratory, Alabama, USA [346] for exome capture and sequencing. Exome capture was performed using the Agilent SureSelect Human All Exon 50 Mb kit (Agilent Technologies, California, USA) and 50 bp paired-end sequencing was subsequently performed on the Illumina HiSeq2000 (Illumina, California, USA) (refer to Appendix 4 for details).

5.3.3 Bioinformatics pipeline for exome sequence analysis

The generated exome data for the eleven FES individuals was analysed with the use of a bioinformatics pipeline, which made use of the following programs: (i) BWA [325] for sequence read alignment; (ii) SAMtools [347] for sorting and indexing of reads; (iii) the GATK [326], including variant quality score recalibration, for variant calling; and (iv) SeattleSeq Annotation 134 [348] for variant annotation. These analyses are detailed in Appendices 4 and 5. Only variants that were assigned a "PASS" value in the "FILTER" field of the vcf files that was generated by GATK were included in downstream analyses. All novel SNPs were submitted to dbSNP [320] and submitted SNP numbers were assigned.

5.3.4 Comparison of the patterns of variation observed in the non-responder and responder individuals

In order to examine the patterns of variation observed in the unrelated non-responder and responder FES individuals, the total and average number of coding variants observed in each group was compared. More specifically, the variants were divided into different classes based on the effect that they were predicted to have on protein function. In each case the percentage of novel variants was investigated. The analyses used to do this are described in Appendix 6.

With regards to the non-responder siblings, all variants that were predicted to abolish the function of the resulting protein product i.e. loss of function (LOF) variants (frameshift, splice-site and stop lost/gained variants) were examined. These LOF variants were then prioritised for the likely role that they play in non-response by excluding (i) all variants that were not shared between the non-responder siblings and (ii) all variants that were present in the responder individuals. This list of LOF variants was further prioritised by identifying variants that occurred in any of the other unrelated non-responder individuals. Allele frequency data was obtained from the 1000 Genomes Project Browser [349].

5.4 Results

Exome sequencing was performed successfully for all eleven FES individuals and the sequence coverage obtained for each individual exome ranged from 61x to 82x, with between 80-87% of the targeted areas obtaining a coverage of at least 30x. A total of 56 346 coding variants were identified in the ten FES individuals, of which 5 557 were novel (9.86%). Comparison of the variation observed in the non-responder and responder FES individuals did not reveal any clear differences, although slightly more coding variants were observed in the non-responder individuals with reference to both the total number of variants (42 678 vs. 42 333 coding variants) and the average number of variants per individual (19 362 vs. 19 253 coding variants) (Table 5.1).

Table 5.1: The total and average number of variants observed in the five non-responder and responder individuals

Class of variation	Non-responders (n=5)		Responders (n=5)	
	Total	Average	Total	Average
Synonymous	22 805	10 509	22 549	10 393
Non-synonymous*	18 917	8 337	18 815	8 336
Frameshift**	652	354	660	363
Splice-site**	204	95	211	95
Stop-gained/lost**	171	66	165	66
Total coding	42 749	19 361	42 400	19 253

* changes the protein product, ** abolishes the function of the protein product

Examination of the novel variants revealed that these variants were far less likely to be shared between the responders and non-responders when compared to known variants, such that 55.63% of known coding variants were shared between the two groups and only 7.43% of novel variants were shared (Figure 5.1). This was particularly the case for the stop gained/lost variants, which are likely to have a large impact on the function of the protein. None of the stop gained/lost variants that were detected in both non-responders and responders were novel. However, a large percentage of the stop gained/lost variants occurring in only the non-responders or the responders were novel (35.71% and 31.25%, respectively). Furthermore, the percentage of novel variants was highly correlated to the predicted effect that the variant has on the protein product, such that LOF variants were more likely to be novel (Figure 5.2). Examination of the variants that were not shared between the non-responders and responders revealed that a large percentage of these variants were present in only one individual, such that 98.00% of the novel variants and 82.17% of the known variants that were unique to either of the groups were present in only one individual.

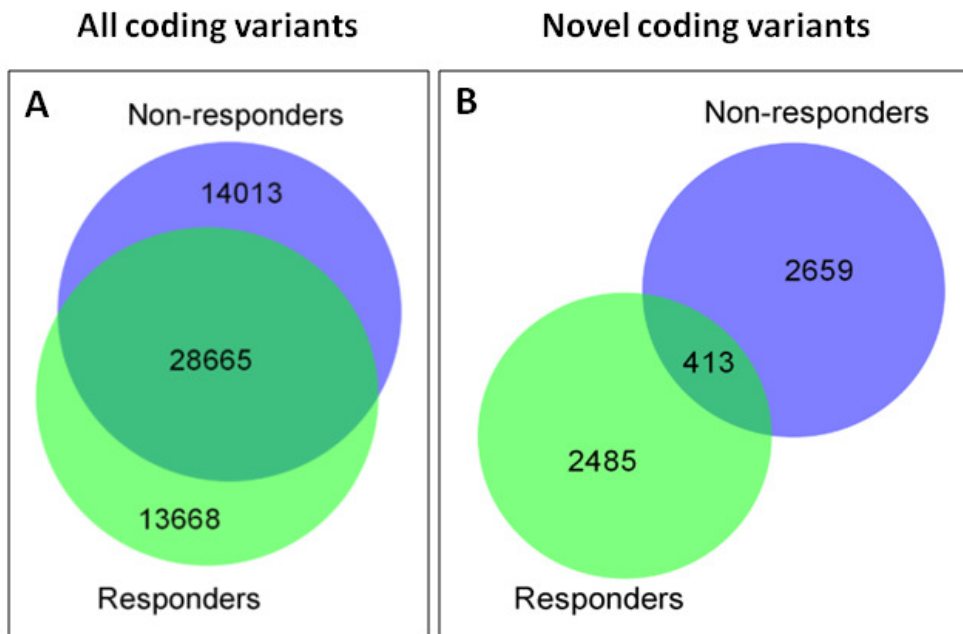


Figure 5.1 Comparison of the coding variation present in the non-responder and responder individuals **A:** The majority of the coding variants that were observed in the ten FES individuals were shared by the non-responders and responders. There were, however, slightly more coding variants observed in the non-responder individuals. **B:** A small portion of the novel coding variants were shared by the two groups, with the majority observed in only the non-responder or responder groups.

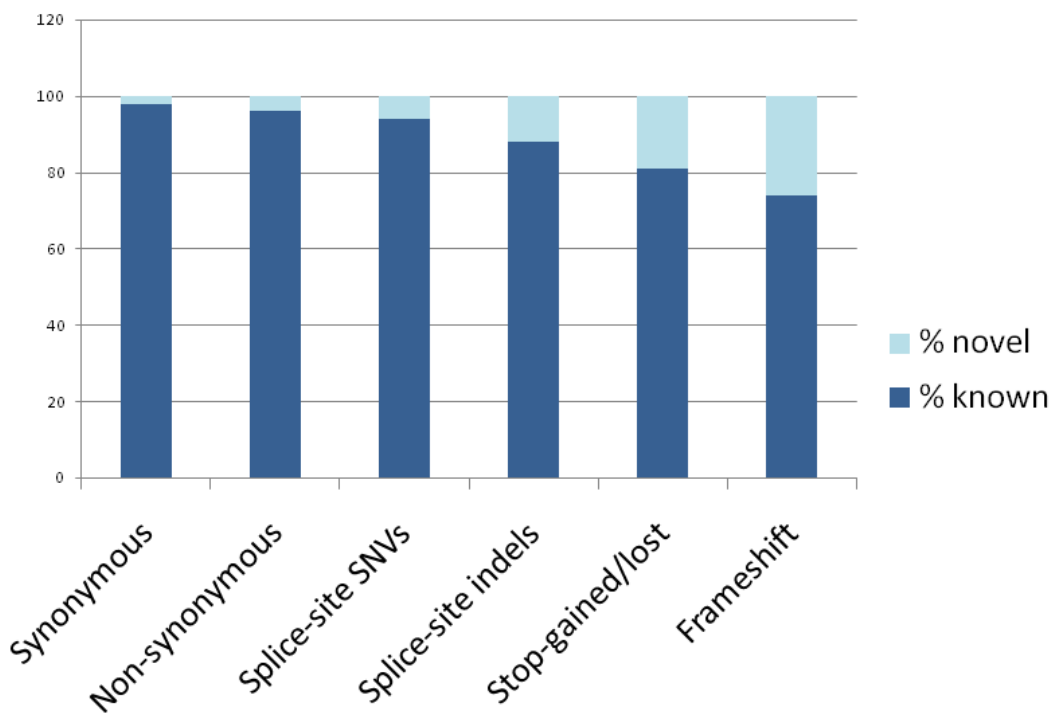


Figure 5.2: The percentage novel variation per individual observed for each class of variant.

When the LOF variants present in the non-responder siblings were examined, it was observed that there were a total of 211 LOF variants that were common to both siblings. However, after removing all variants that were also present in the responder individuals, only 22 LOF variants remained, of which six were present in at least one of the additional unrelated non-responder individuals. When examining all 22 LOF variants it was observed that nine of these variants were novel (ss678319342 in *C6*, ss678371274 in *DNMBP*, ss678329213 in *HIST1H2BM*, ss678372253 in *NRAP*, ss678431152 in *SLC25A41*, ss678374868 and ss678374869 in *STK33*, ss678424277 in *TMEM235*, ss678333610 in *ZBTB24*), there was no population frequency data available on the 1000 Genomes Project Browser for two variants (rs71710115 in *HOMEZ*, rs35744335 in *PKD1L3*) and one variant (rs11368509 in *UPP2*) has only been detected in the Khoisan population to date [280]. The allele frequencies of the remaining ten variants in the African, Asian and European populations genotyped by the 1000 Genomes Project are displayed in Table 5.2. Eight of these ten SNPs have allele frequencies of less than 5% in the both Asian and European populations (refer to shaded areas in Table 5.2).

Table 5.2: The allele frequencies of the loss of function variants present in the non-responder siblings, but absent from the responder individuals

Variant	Gene	LOF variant	AFR	ASN	EUR
rs112899189	<i>CLU10S</i>	Frameshift	0.03	0.00	0.01
rs10666583	<i>GRIN3B</i>	Frameshift	0.07	0.08	0.28
rs73439094	<i>C6orf52</i>	Splice-5	0.26	0.03	0.00
rs8065203	<i>CYTH1</i>	Splice-3	0.03	0.00	0.00
rs57118523	<i>HK1</i>	Splice-3	0.07	0.00	0.00
rs149764161	<i>MYL5</i>	Splice-3	0.01	0.00	0.00
rs17104991	<i>SLC25A21</i>	Splice-5	0.12	0.00	0.00
rs74141230	<i>TRIM17</i>	Splice-3	0.15	0.01	0.00
rs28759013	<i>TXNDC16</i>	Splice-3	0.11	0.00	0.00
rs80220955	<i>OR5AC2</i>	Stop-lost	0.08	0.01	0.23

AFR: African 1000 Genomes Project population, ASN: Asian 1000 Genomes Project population, EUR: European 1000 Genomes Project population, LOF: Loss of function. Variants with minor allele frequencies < 0.05 are shaded in grey.

5.5 Discussion

To our knowledge, this is the first published study utilising exome sequencing to examine the pharmacogenomics of antipsychotic response. Furthermore, to date there is only one publically available non-European South African genome sequence, namely that of Archbishop Desmond Tutu [280]. Thus, this study provides important data regarding the variation present in South African schizophrenia patients. This is of significance as African populations are underrepresented in both genomic and psychiatric research [240,241].

When comparing the data generated by this project to the 'healthy' high coverage trio genomes sequenced by the 1000 Genomes Project [281], it was observed that the number of coding variants observed in the FES individuals fell within the individual range reported by the 1000 Genomes Project. Comparison of the non-responders and responders revealed

that there were slightly more variants present in the non-responder individuals, although these differences were minimal. Of greater significance for future studies was the identification of LOF variants that were shared between the sibling non-responders but were not present in the responder individuals, the significance of which should be investigated in a larger cohort. These analyses identified a total of 22 variants in 21 genes, of which none were present in all of the non-responder individuals. The inability of this study to identify a single variant that could completely explain non-response, even with the use of well characterised patients, provides further evidence that antipsychotic treatment response is a complex trait. This draws attention to the likelihood that different genes and variants may be involved in treatment response outcomes in different individuals. Thus, although the gene networks and pathways that are disrupted may be common to all non-responder individuals, the specific variants present may differ between individuals.

To further examine the role that the 21 genes that were disrupted in the non-responder individuals may play in antipsychotic response, a literature search was performed in the PubMed Database [82] utilising each of the gene names and the term “antipsychotic”. The results from this search revealed that only one of the genes, *GRIN3B*, had a previous connection to antipsychotic response [350]. The lack of information pertaining to the role that these genes play in antipsychotic response is likely related to the fact that very little is known regarding the function of the majority of these genes, as recorded on NCBI’s gene resource [333]. Thus, these genes are unlikely to have been included in past candidate gene studies. With reference to the specific variants, of the 22 prioritised LOF variants, 20 were novel or had $MAF < 0.05$ in the 1000 Genomes Project Asian and European populations. Although, to our knowledge, none of the 22 variants have been reported in past antipsychotic research, all but two occur at very low frequencies in the Asian and European populations. These populations have been the focus of the majority of such studies, leaving African populations under-represented in antipsychotic pharmacogenomic studies [244]. Thus, variants affecting African individuals may differ to those affecting non-African individuals. In addition, the fact that many of these LOF variants were either novel or occurred at low frequencies highlights the likely role that rare or novel variants may play in the non-response phenotype.

The role that these variants play in non-response was further highlighted when examining the global patterns of novel variation, with particular reference to the LOF variants. It was observed that there were a large number of novel variants that were unique to either the non-responders or responders (Figure 5.1) and 98.00% of this unshared novel variation was only observed in one individual. Therefore, it is likely that this variation may be rare, although genotyping of these variants in a larger cohort will be necessary to confirm this. Furthermore, when examining the different classes of variation, it was observed that the LOF class showed the highest percentage of novel variants (Figure 5.2). A recent publication examining autism has reported a two-fold increase in rare LOF variants ($MAF < 0.05$) in cases when compared to controls [351]. As the genetics of schizophrenia and autism have been shown to overlap [352], such findings may be transferable to schizophrenia and related phenotypes. Due to the fact that examination of known common variation through the use of GWAS has yielded incomplete information regarding the contribution of genetics to schizophrenia and antipsychotic treatment response, it remains likely that the some of the answers to the missing heritability lie in rare and novel variation, which can only be detected through sequencing.

Sequencing studies examining schizophrenia have also emphasised the potential role that rare and *de novo* variants play in the development of schizophrenia [65–68]. As these variants are unlikely to be shared between different populations [71], it may be necessary to examine many different populations to determine the exact variation contributing to poor treatment outcomes and the development of schizophrenia. This is especially relevant for southern African populations, which exhibit high levels of variation, but have been under-represented in genomic research to date [240,241,280].

5.6 Conclusions

This study utilised well characterised patients on extreme ends of the antipsychotic treatment spectrum, as well as the addition of sibling non-responders, to prioritise variation that may be involved in antipsychotic treatment. These analyses identified 22 LOF variants that may be involved in antipsychotic non-response, the effects of which should be investigated in a larger cohort. It should also be noted that the use of schizophrenia patients who have been treated for several years and have never responded to treatment, rather than FES patients, may have increased the power of this study. Nonetheless, the results from this study have highlighted the complexity of antipsychotic treatment response, as well as the importance of rare and novel LOF variation in this phenotype. This reiterates the need for WGES to detect the entire spectrum of variation, particularly in the under-represented African populations. Furthermore, these genotyping strategies should be accompanied with statistical methods to account for variation in entire pathways, the importance of which has been highlighted by Kiezun *et al.* [339].

5.7 Future perspectives

In the next five-ten years, WGES will become more affordable and more accessible. Past psychiatric research has focused predominantly on common variants; however, it seems likely that it may become more important to focus on rare variants. It appears that the variants involved in antipsychotic treatment response may differ between individuals; however, the pathways involved may be common. Thus, in future, statistical methods that are able to examine rare variants in multiple genes will be necessary. The implementation of these strategies may identify novel drug targets and aid in optimal antipsychotic treatment regimes.

Financial disclosure/acknowledgements:

The work reported here was supported by the following grants to the authors : BI Drögemöller: NRF research bursary and the L'Oréal-UNESCO for women in Science in Sub-Saharan Africa Fellowship; DJH Niehaus: South African MRC operating research grant; B Chiliza: South African MRC and Hamilton Naki Clinical Research Fellowship; L Asmal: None; AK Malhotra: acts as a consultant for Genomind Inc and has received a grant from AbbVie; GEB Wright: None; R Emsley: has participated in speakers/advisory boards and received honoraria from AstraZeneca, Bristol-Myers Squibb, Janssen, Lilly, Lundbeck, Organon, Pfizer, Servier, Otsuka and Wyeth. He has received research funding from Janssen, Lundbeck and AstraZeneca. This study was supported by a NEPAD grant, through the DST of South Africa; L Warnich: NRF operating research grant. This financial assistance is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily attributed to these funding sources.

Declaration by the candidate:

With regard to Chapter 5 (p 67-74), the nature and scope of my contribution were as follows:

Nature of contribution	Extent of contribution (%)
Involved in grant applications Design of project Identification of suitable exome sequencing service provider gDNA extraction gDNA quality control Preparation of gDNA for shipping Transfer of generated sequence data to high performance computer Regular backing up of generated data Sorting and indexing of sequence reads Read alignment recalibration Variant calling Variant annotation Submission of novel variants to dbSNP Assessment of the type and number of variants in each individual Comparison of the variants present in the non-responder and responder individuals Prioritisation of variants shared between sibling non-responders 1000 Genomes Project comparisons Interpretation of data Writing of article	75%

The following co-authors have contributed to Chapter 5 (p 67-74):

Name	Email address	Nature of contribution	Extent of contribution (%)
Prof Louise Warnich	lw@sun.ac.za	Funding and design of project Identification of suitable exome sequencing service provider	12.5%
Dr Galen Wright	galen@sanbi.ac.za	Interpretation of data Critical assessment of article	
Prof Robin Emsley	rae@sun.ac.za	Funding	12.5%
Prof Dana Niehaus	djhn@sun.ac.za	Design of project Patient recruitment	
Dr Bonginkosi Chiliza	bonga@sun.ac.za	Patient diagnosis	
Dr Laila Asmal	laila@sun.ac.za	Patient treatment	
Dr Anil Malhotra	AMalhotra@nshs.edu	Clinical assessment of patients Critical assessment of article	

CHAPTER 6: The identification of novel genetic variants and pathways affecting antipsychotic treatment response in South African first episode schizophrenia patients

Drögemöller Britt I¹, Chiliza Bonginkosi², van der Merwe Lize^{3,4}, Wright Galen EB^{1,5}, Daya Michelle⁴, Hoal Eileen⁴, Asmal Laila², Niehaus Dana JH², Emsley Robin², Warnich Louise¹

¹Stellenbosch University, Department of Genetics, Stellenbosch, South Africa;

²Stellenbosch University, Department of Psychiatry, Tygerberg, South Africa;

³University of the Western Cape, Department of Statistics, Bellville, South Africa;

⁴Stellenbosch University, Department of Molecular Biology and Human Genetics, Tygerberg, South Africa

⁵University of the Western Cape, South African Bioinformatics Institute, Bellville, South Africa

6.1 Summary

Introduction: Although the administration of antipsychotics is an integral part of schizophrenia treatment, the use of these agents is not equally effective in all patients. Therefore the field of pharmacogenomics may play a valuable role in the optimisation of antipsychotic treatment. However, in order for pharmacogenomic studies to be successful, careful study designs utilising extensive clinical and genomic data need to be implemented.

Materials and methods: This study utilised a well characterised cohort of first episode schizophrenia patients who were treated with flupenthixol decanoate (depot injections) for twelve months during which the response to treatment was regularly assessed. The exomes of a subset of patients on extreme ends of the treatment response spectrum were sequenced. These data were used in combination with the results from previous antipsychotic studies to design a panel of variants for genotyping in the entire cohort. The genetic data were used to identify associations with treatment response outcomes. **Results:** Eleven variants were significantly associated with treatment response outcomes ($P < 2.19 \times 10^{-5}$). These included two variants that have been significantly associated with antipsychotic pharmacogenetic traits in past GWAS and nine variants which were predicted to change the function of the corresponding proteins, eight of which were novel. Examination of the function of these proteins revealed a potential role for (i) folate metabolism with regards to the positive symptom domain and (ii) proteins which may be involved in neuronal migration with regards to the negative symptom domain. **Conclusion:** This study has demonstrated the value of well characterised cohorts and genomic data for antipsychotic pharmacogenomic applications. The use of these strategies made it possible to identify novel genetic variants that may be involved in antipsychotic treatment response, as well as to replicate past findings with regards to antipsychotic pharmacogenomics. These findings should play a role in improving our understanding of antipsychotic treatment response both in the context of South Africa and globally.

Keywords:

Antipsychotics, exome sequencing, first episode schizophrenia, treatment response, South Africa

6.2 Introduction

As detailed in Chapters 2.1.1 and 2.2.2, schizophrenia is a poorly understood and complex disorder that occurs the world over [9,10], the symptoms of which are most effectively treated with the use of antipsychotics [15]. Nonetheless, treatment outcomes associated with antipsychotics are non-optimal, unpredictable and vary widely between patients with only approximately 60% of patients responding to antipsychotic treatment [16,84,86]. Thus, methods to optimise antipsychotic treatment are urgently required.

It is for this reason that there is much anticipation for the implementation of antipsychotic pharmacogenomics. If genetic variants can be identified that will predict the optimal antipsychotic treatment required for each individual prior to dosing, the occurrence of ADRs and non-response can be reduced and the long term outcomes of schizophrenia patients can be improved (Figure 2.7). As highlighted in Chapter 4, there have been many studies examining antipsychotic pharmacogenomics; however, they have focused predominantly on candidate genes (Table S2). The results obtained from these studies have been inconsistent and the identified pharmacogenetic variants have been shown to have little clinical relevance [176,195].

The lack of clinically relevant results obtained from antipsychotic pharmacogenetic studies could in part be attributed to the fact that the mechanism of action of antipsychotics is only partially understood [96]. By replacing candidate gene approaches with strategies that scan the entire genome, we may be able to improve our understanding of antipsychotic pharmacogenomics and identify novel pathways involved in treatment response to this class of medications [241]. Unfortunately, to date there are only a handful of studies that have utilised GWAS for antipsychotic pharmacogenomics and these studies have used different cohorts and examined different drugs and phenotypes [203,204,206–214] (Table 2.3).

An interesting observation from the past antipsychotic pharmacogenetic studies, both candidate gene studies and GWAS, is that significant associations have been detected with several different genes and variants (Table S2 and Table 2.3). This highlights the complexity of antipsychotic treatment response and it seems likely that many variants, genes and pathways may be involved in influencing antipsychotic treatment outcomes. In addition to this, the antipsychotic treatment response phenotype has been reported to be a non-uniform trait with patients exhibiting different treatment response trajectories [103]. Further complicating analyses is the fact that the symptoms observed in schizophrenia patients vary from patient to patient, with regards to both the type of symptoms and the severity of these symptoms [36]. Thus, in order to identify the genetic variants contributing to antipsychotic treatment response, patients will need to be carefully characterised in terms of clinical and genomic information and antipsychotic treatment response will need to be assessed with reference to the different symptom domains.

The value of well characterised cohorts in combination with extensive genomic data is highlighted when examining the antipsychotic GWAS that were described in Table 2.3. Interestingly, of these eleven GWAS, the most significant association ($P=6 \times 10^{-12}$) was obtained in a study by Malhotra *et al.* [203], which utilised a well characterised FES cohort to identify an association with weight gain. The significant association detected by this study highlights two key points in antipsychotic pharmacogenomic research, namely the

importance of (i) genome-wide technologies for the identification of novel variants related to treatment response outcomes and (ii) the use of well characterised cohorts of patients. Although the significance of well characterised patients was detailed in Chapter 2.3.3, in summary the four most important aspects that need to be considered for antipsychotic pharmacogenetics are related to (i) eliminating differences in disease progression, (ii) standardised treatments, (iii) compliance and (iv) longitudinal assessments. [176,180,195,217].

Another important aspect of antipsychotic pharmacogenomics that has been largely overlooked is the lack of data pertaining to African populations. Not only are there no antipsychotic GWAS performed in Africa [202], but research pertaining to candidate genes is also lacking in these populations [240,244]. This is of serious consequence, as was brought to attention in Chapter 1, due to the fact that the burden of schizophrenia, measured in DALYs, in LMIC is ten times higher than in high-income countries [12]. Thus, optimal treatment is urgently required. In addition, as described in Chapter 3, there is a lack of genomic information available for populations residing in Africa. Therefore antipsychotic pharmacogenomic studies using WGES are necessary in Africa.

Unfortunately, the lack of resources in Africa may hamper the implementation of large scale genomic studies in the context of South Africa [4]. Thus, innovative strategies that allow for the identification of risk alleles from several populations and provide information regarding the genetic factors contributing to antipsychotic treatment response, without excessive costs, are required. These strategies may include the use of the SAC population, which has ancestry contributions from several different sources and can thus be used to detect risk alleles from different populations [242]. In addition, the careful characterisation of patients allows for the identification and subsequent analysis of individuals on extreme ends of the phenotypic spectrum. WGES of this subset of patients can provide information regarding the genetic variation with the greatest influence on the phenotype of interest [353]. Lastly, as WGES will detect several variants, the use of family members can help to reduce the noise created by non-causal variants [353]. Therefore, using a combination of innovative strategies, the costs associated with genomic technologies can be reduced and valuable information can be obtained.

6.3 Materials and Methods

6.3.1 Patient samples

The patients described in Chapter 5 were utilised and venous blood samples were obtained for 104 of these patients [gender: 79% male; median age (years): 24±7; ethnicity: 80% SAC, 13% Xhosa and 8% European descent]. All gDNA samples were extracted from the venous blood samples by means of the Miller *et al.* [354] protocol. As described previously, patients received flexible doses of flupenthixol decanoate over twelve months and were regularly assessed by means of the PANSS (Table 6.1). In addition, patients were assessed with regards to treatment refractoriness over the twelve months. Treatment refractoriness was defined as: (i) study discontinuation due to poor response, (ii) endpoint percentage change in PANSS total score <20% or (iii) endpoint PANSS total score >70, provided that patients meeting criteria (ii) and (iii) had completed at least three months of treatment and had not

experienced a relapse. A total of nine patients met these criteria. Extensive demographic and medical information was also collected for each patient at the baseline assessments.

Table 6.1: PANSS score assessments at baseline and twelve months

	Median	Lower Quartile	Upper Quartile
Baseline PANSS-P score	24	22	27
Month 12 PANSS-P score	8	7	9
Baseline PANSS-N score	25	21	29
Month 12 PANSS-N score	12	10	17
Baseline PANSS-G score	45	40	51
Month 12 PANSS-G score	21	18	26
Baseline PANSS-T score	94	85	105
Month 12 PANSS-T score	43	37	52

PANSS: Positive and Negative Syndrome Scale, P: Positive, N: Negative, G: General pathological, T: Total

6.3.2 Prioritisation of variants for genotyping in the entire FES cohort

Patients falling on extreme ends of the treatment response spectrum were selected for exome sequencing and these samples were sequenced and analysed as previously described in Chapter 5. These exome data were used in combination with the literature to prioritise variants for genotyping in the entire FES cohort. The variants prioritised for genotyping included: (i) the three-six variants that were most significantly associated with antipsychotic treatment outcomes in each of the past GWAS; (ii) variants that have most frequently been reported to be associated with antipsychotic treatment outcomes in past candidate gene studies as reported in the review articles described in Chapter 4; (iii) variants in the top 25 candidate antipsychotic pharmacogenes (Table 2.2) that were identified in the exome sequence data and were predicted to alter the function of the genes; and (iv) variants present in the exome data that were identified from variant and gene based analyses performed using the Variant Annotation, Analysis and Search Tool (VAAST) [355] as described by the software authors. The VAAST analyses removed all variants that were not common to the sibling non-responders and used different analyses to model the inheritance mode of antipsychotic response. In addition, VAAST was used to identify variants that occurred in three or more of the non-responders and none of the responders (and *vice versa*) and were predicted to alter the function of the genes in which they occurred. The predicted effect of variants on gene function was assessed with the use of SIFT [308], PolyPhen-2 [356] and SeattleSeqAnnotation134 [348] (refer to Appendix 7 for more details regarding the selection process).

The prioritised list of variants was subsequently assessed with Illumina's Assay Design Tool (ADT) to measure the feasibility of genotyping each of the SNPs using a 384-plex BeadXpress Assay (Illumina, California, USA). Where possible, all variants that received failure codes from this tool were tagged with variants identified from the exome and/or 1000 Genomes Project data ($r^2 > 0.8$) (refer to Appendix 7 for specific details). In cases where no tagSNPs were available for these failed variants, PCR-RFLP/TaqMan assays were designed instead. Employing these criteria, 293 variants were prioritised for genotyping in the entire cohort. Furthermore, 100 AIMs were included for genotyping in the entire cohort. These AIMs were

selected from genome-wide SNP array data from the five populations that have been reported to contribute to the ancestry of the SAC population (African San, African non-San, European, South Asian and East Asian) as described by Daya *et al.* [357]. A summary of the variants that were prioritised for genotyping in the entire cohort is displayed in Figure 6.1 and more detailed descriptions of the strategies employed are provided in Appendix 7.

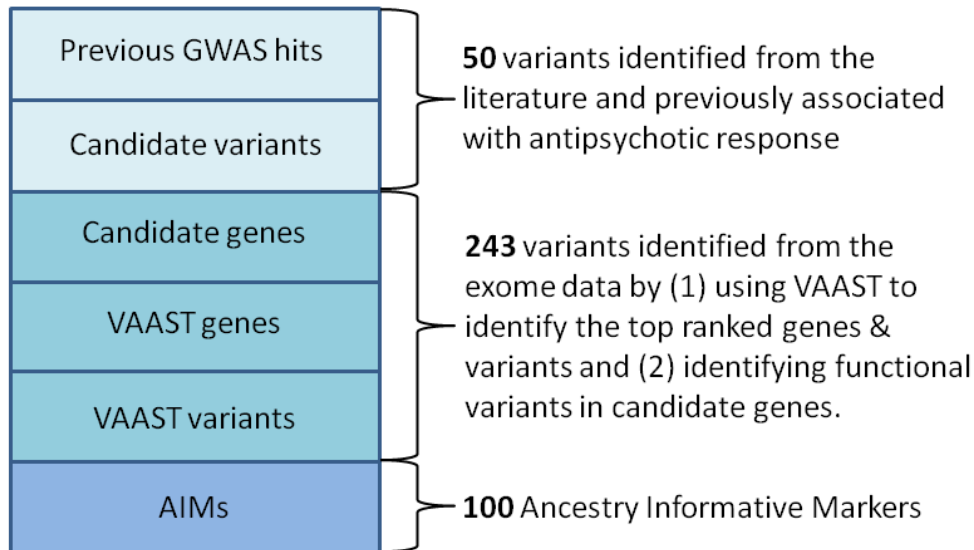


Figure 6.1: The selection of variants for genotyping using the BeadXpress Assay.

6.3.3 Genotyping of the prioritised variants in the entire FES cohort

Genotyping of the variants that passed the Illumina ADT assessment was performed at the University of Utah Genomics Core Facility [358] with a 384-plex Illumina BeadXpress Assay using VeraCode technology. Duplicate samples were included as internal controls. Thereafter, genotype clustering of the variants was performed using default setting on the GenomeStudio Data Analysis Software (Illumina, California, USA). Samples were examined for exclusion based on the DNA report or incorrect gender assignment (as determined by heterozygosity for X chromosome variants) and genotype concordance between the exome and BeadXpress data was examined. All clusters were manually inspected and variants with unsatisfactory genotype clusters or call rates less than 90% were excluded (refer to Appendix 8 for more details). For the nine variants that failed the ADT analysis and could not be tagged by other variants, PCR-RFLP assays were designed for eight of these variants and a custom TaqMan assay was designed for the remaining variant. Details regarding the design of these assays and the specific genotyping conditions used are provided in Appendix 8.

6.3.4 Statistical analyses

Allele and genotype frequencies, as well as deviations from Hardy–Weinberg equilibrium (HWE), were determined for all genotyped polymorphisms. HWE analyses were performed in Tools For Population Genetic Analysis (TFPGA) Software v1.3 [359] using a chi-squared or Fisher’s exact test. $P < 0.002$ were considered significant for HWE testing. Only variants that were successfully genotyped or had $MAF > 0.01$ were included in the subsequent analyses of the unrelated individuals. To identify allelic and genotypic associations with the twelve months longitudinal PANSS scores, a mixed-effects model repeated measures analysis was utilised. Due to the skewed distributions of these scores, log transformations were

performed. For the treatment refractory analyses, a logistic regression model was used. To allow for the correction of population stratification, ancestry proportions were estimated by ADMIXTURE [360] utilising the AIMs data obtained for the FES cohort, as well as the data obtained from the five populations from which the AIMS were designed. All association analyses were subsequently adjusted for age, gender and the proportions of ancestry of the different groups for each individual calculated by ADMIXTURE analyses. In addition, the mixed-effects model repeated measures analyses were adjusted for baseline PANSS scores. Inheritance models were tested for all significant allelic and genetic associations and the most significant of the models are reported in the results ($P < 2.19 \times 10^{-5}$). Effect estimates with 95% confidence intervals are derived from the models. Graphs of observed values are presented as an indication of the corresponding unadjusted patterns in the data. All statistical analyses were performed in R [336] using the R package, genetics (version 1.3.6) [361].

6.4 Results

After exclusion of all variants that failed quality control, a total of 347 variants (252 prioritised variants and 95 AIMs) remained (90.36% success rate). Removal of all variants with $MAF < 0.01$ resulted in the inclusion of 228 variants for the subsequent statistical analyses. Examination of these variants in the FES individuals revealed that all patients were successfully genotyped and duplicate samples correlated. Thus, all unrelated individuals ($n=103$) were included in the subsequent statistical analyses. All variants on autosomal chromosomes that were included in the downstream analyses were in HWE, with the exception of rs4926044, which was included in the subsequent analyses but was flagged. Furthermore, the genotypes obtained from the exome data for these variants were compared to the BeadXpress data. In cases where the genotypes called from the two strategies did not correlate, GenomeStudio was used to manually inspect the clusters generated by the BeadXpress Assay, while SAMtools tview [347] was used to manually inspect the alignments of the exome sequence reads. Any lack of correlation between these genotyping strategies could be attributed to low coverage of the exome sequence data, misalignment of the sequence reads or incorrect BeadXpress calling. In the case of incorrect BeadXpress calling, the clusters were either redefined according to the correct genotype or the clusters were tightened to eliminate incorrect genotype assignment. Two SNPs (rs11556167 and rs150402481) did not correlate between the genotyping strategies and this could not be attributed to poor exome or BeadXpress genotyping. Therefore these two SNPs were flagged.

Statistical analyses of the 228 successfully genotyped variants revealed that, after Bonferroni correction for multiple testing, eleven variants were significantly associated with the five traits that were examined with regards to the genotypic and allelic models that were utilised ($P < 2.19 \times 10^{-5}$). All variants, with the exception of rs2027937 ($OR_{G-allele}=161.4$, 95% CI 12.1–7954.3, $P=1.0 \times 10^{-5}$) which was associated with treatment refractoriness (Figure 6.2), were associated with change in PANSS scores over twelve months of treatment with flupenthixol decanoate. Closer inspection revealed that eight of these variants occurred in novel loci that were identified from the VAAST exome analyses. All of these variants were predicted by SIFT and/or PolyPhen-2 to alter the function of the protein, with the exception of one variant which was used to tag a missense variant (rs10153210) in *TCF25* ($r^2=1$,

LOD=3.3). With regards to the remaining three variants, two have been associated with antipsychotic treatment response outcomes in past GWAS and one has been associated with antipsychotic treatment response outcomes in past candidate gene studies. The effect sizes and inheritance models of these variants, as well as the predicted effects that the variants have on their corresponding protein products, are displayed in Table 6.2, Table 6.3 and Figure 6.3, as well as Appendix 10.

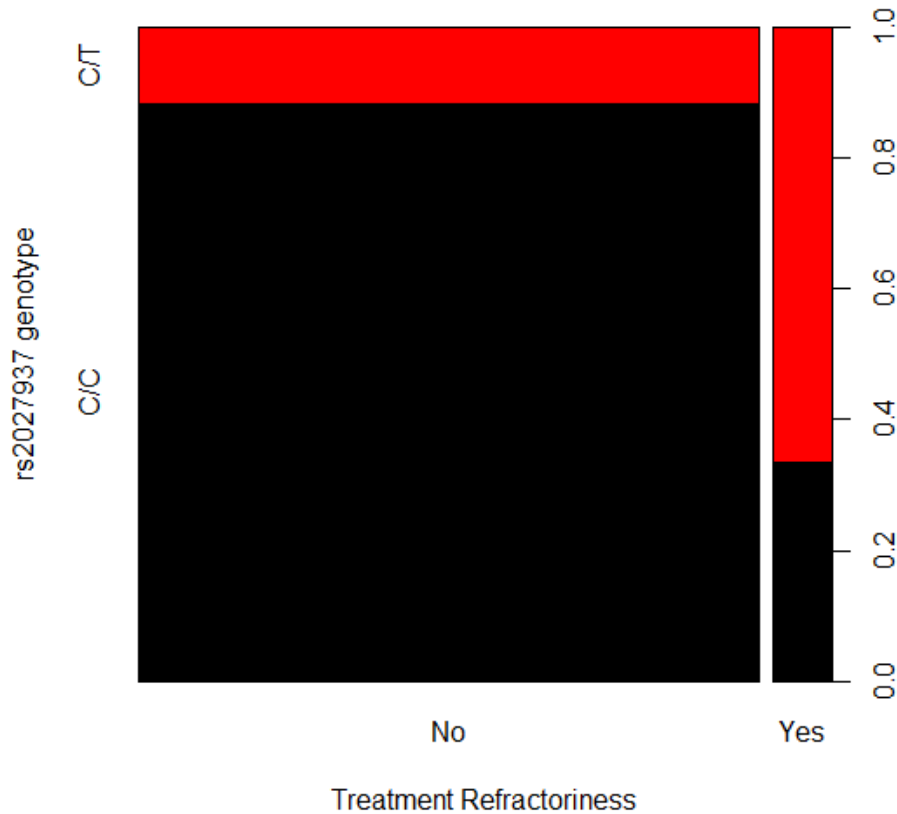


Figure 6.2: A mosaic plot representing the observed genotype frequencies for rs2027937 associated with treatment refractoriness, with the width of the plots representing the number of individuals in each group. Those individual who were not treatment refractory (n=94) exhibited at MAF=0.05, while those who were treatment refractory (n=9) exhibited a MAF=0.33.

Table 6.2: The effect sizes (estimated difference in slope) and corresponding 95% confidence intervals of the significantly associated variants for specific inheritance models on the traits

Locus name	Gene	Trait	P value	Model	PANSS Slope Difference	95% CI	
rs3924426	<i>SLCO3A1</i>	PANSS-N score	9.6x10 ⁻⁷	Recessive	CC vs. TT+TC	-6.0%	-8.1% -3.8%
		PANSS-T score	4.7x10 ⁻⁶			-4.9%	-6.7% -3.0%
rs1801133	<i>MTHFR</i>	PANSS-P score	2.1x10 ⁻⁶	Additive	AA vs. AG vs. GG	3.6%	2.1% 5.1%
		PANSS-T score	1.2x10 ⁻⁵			2.3%	1.3% 3.4%
rs10380	<i>MTRR</i>	PANSS-P score	3.1x10 ⁻⁶	Dominant	TT+CT vs. CC	-3.0%	-4.4% -1.6%
rs36062234	<i>DNAAF1</i>	PANSS-N score	3.1x10 ⁻⁶	Recessive	GG vs. CG+CC	-10.0%	-13.9% -6.0%
rs112033334	FREM3	PANSS-G score	4.2x10 ⁻⁶	Recessive	AA vs. AG+GG	-10.7%	-15.0% -6.2%
		PANSS-T score	3.1x10 ⁻⁵			-8.8%	-12.9% -4.5%
rs7245949	<i>MUC16</i>	PANSS-T score	4.9x10 ⁻⁶	Dominant	AA+AG vs. GG	-2.2%	-3.1% -1.2%
		PANSS-N score	1.6x10 ⁻⁵			-2.6%	-3.7% -1.5%
rs10923472	<i>SPAG17</i>	PANSS-N score	7.3x10 ⁻⁶	Additive	AA vs. AG vs. GG	2.0%	1.1% 2.9%
rs10153210	<i>TCF25</i>	PANSS-N score	1.2x10 ⁻⁵	Additive	CC vs. CT vs. TT	-2.8%	-4.0% -1.6%
rs10805321	Intergenic	PANSS-G score	1.8x10 ⁻⁵	Recessive	AA vs. CC+CA	-5.5%	-8.2% -2.7%
rs149570530	<i>DNHD1</i>	PANSS-N score	1.8x10 ⁻⁵	NA*	CG vs. CC	-6.2%	-9.0% -3.3%
		PANSS-G score	4.7x10 ⁻⁵			-5.9%	-8.5% -3.3%

*As there were no homozygotes for the variant, the model of inheritance could not be conclusively determined. PANSS: Positive and Negative Syndrome Scale, P: Positive, N: Negative, G: General pathological, T: Total, NA: Not applicable.

Table 6.3: The significantly associated variants and their predicted effects on traits

Locus name	Trait	Gene	Effect on gene product	PolyPhen-2	SIFT
rs3924426 ¹	PANSS-N score PANSS-T score	<i>SLCO3A1</i>	Intronic	NA	NA
rs1801133 ²	PANSS-P score PANSS-T score	<i>MTHFR</i>	A222V	Probably damaging	Damaging
rs10380	PANSS-P score	<i>MTRR</i>	H622Y	Benign	Damaging
rs36062234	PANSS-N score	<i>DNAAF1</i>	D387E	Possibly damaging	Tolerated
rs112033334	PANNS-G score PANSS-T score	<i>FREM3</i>	H1500Y	Not predicted	Damaging
rs7245949	PANSS-T score PANSS-N score	<i>MUC16</i>	T2891I	Probably damaging	Not predicted
rs10923472	PANSS-N score	<i>SPAG17</i>	P1348L	Probably damaging	Damaging
rs10153210 ³	PANSS-N score	<i>TCF25</i>	Intronic	NA	NA
rs10805321 ¹	PANSS-G score	Intergenic	NA	NA	NA
rs149570530	PANSS-G score PANSS-N score	<i>DNHD1</i>	F1360L	Probably damaging	Damaging
rs2027937	Treatment refractoriness	<i>CCHCR1</i>	A367T	Probably damaging	Tolerated

¹: Previous GWAS association, ²: Previous candidate gene association, ³: In LD ($r^2=1$, LOD=3.33 in exome data) with rs13338056 which results in a K53Q amino acid change in *TCF25*, PANSS: Positive and Negative Syndrome Scale, P: Positive, N: Negative, G: General pathological, T: Total, NA: Not applicable

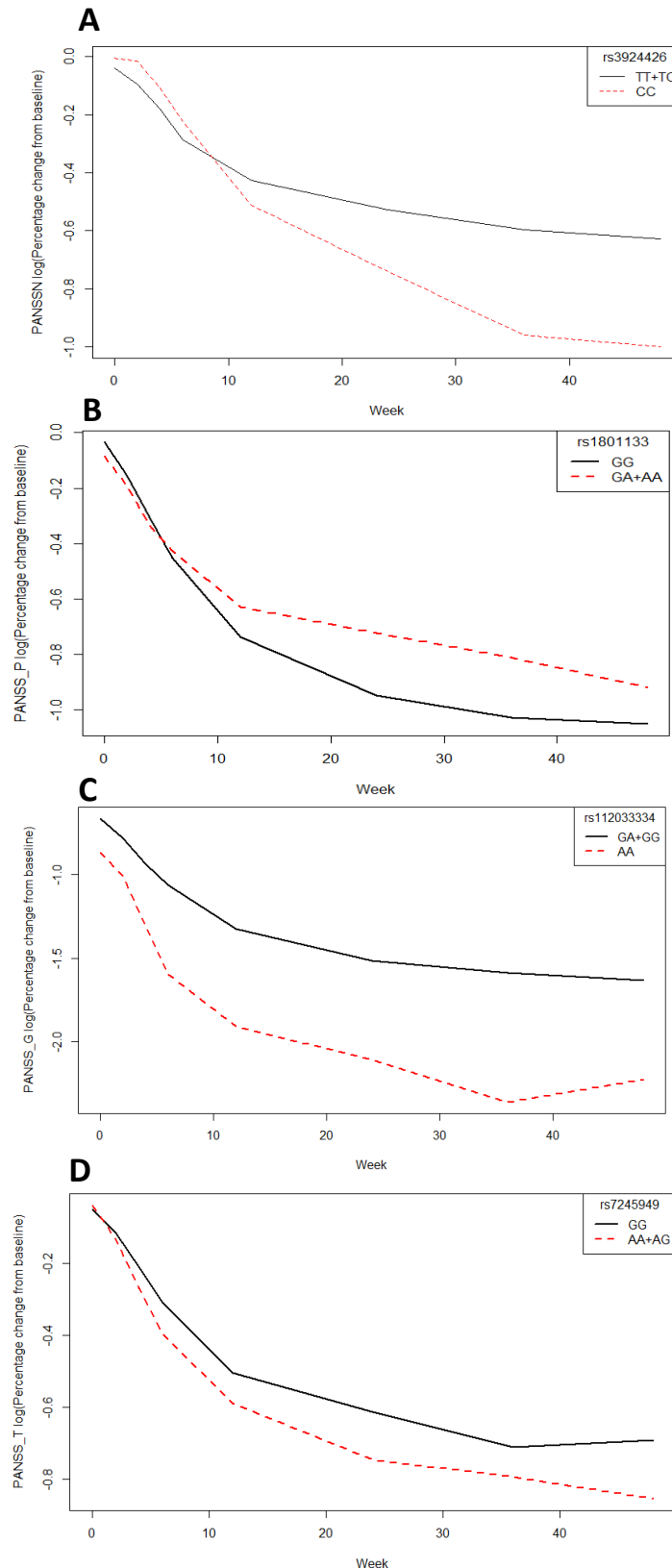


Figure 6.3: The differences in the patterns of PANSS scores for the variants with the most significant P values for each of the symptom domains. **A:** Association of rs3924426 with PANSS negative scores, **B:** Association of rs1801133 with PANSS positive scores, **C:** Association of rs112033334 with PANSS general pathological scores, **D:** Association of rs7245949 with PANSS total scores.

Examination of the five ancestral contributions to the SAC FES individuals revealed that the proportion of ancestry present in each individual differed substantially across the cohort (Table 6.4 and Figure 6.4). Due to the fact that the SAC population comprised the majority of the FES cohort (n=82), the allele frequencies of the eleven significantly associated variants in this highly admixed population were compared to those reported for the three 1000 Genomes Project populations that have contributed most substantially to the ancestry of the SAC population, namely the African (AFR), Asian (ASN) and European (EUR) descent populations (Figure 6.5).

Table 6.4: Maximum, minimum and median proportions of ancestry observed in the SAC FES individuals

	African (Non-San)	African (San)	European	East Asian	South Asian
Minimum	0.00001	0.00001	0.00001	0.00001	0.00001
Median	0.26287	0.22623	0.22781	0.13308	0.12769
Maximum	0.70783	0.59357	0.55550	0.60614	0.52522

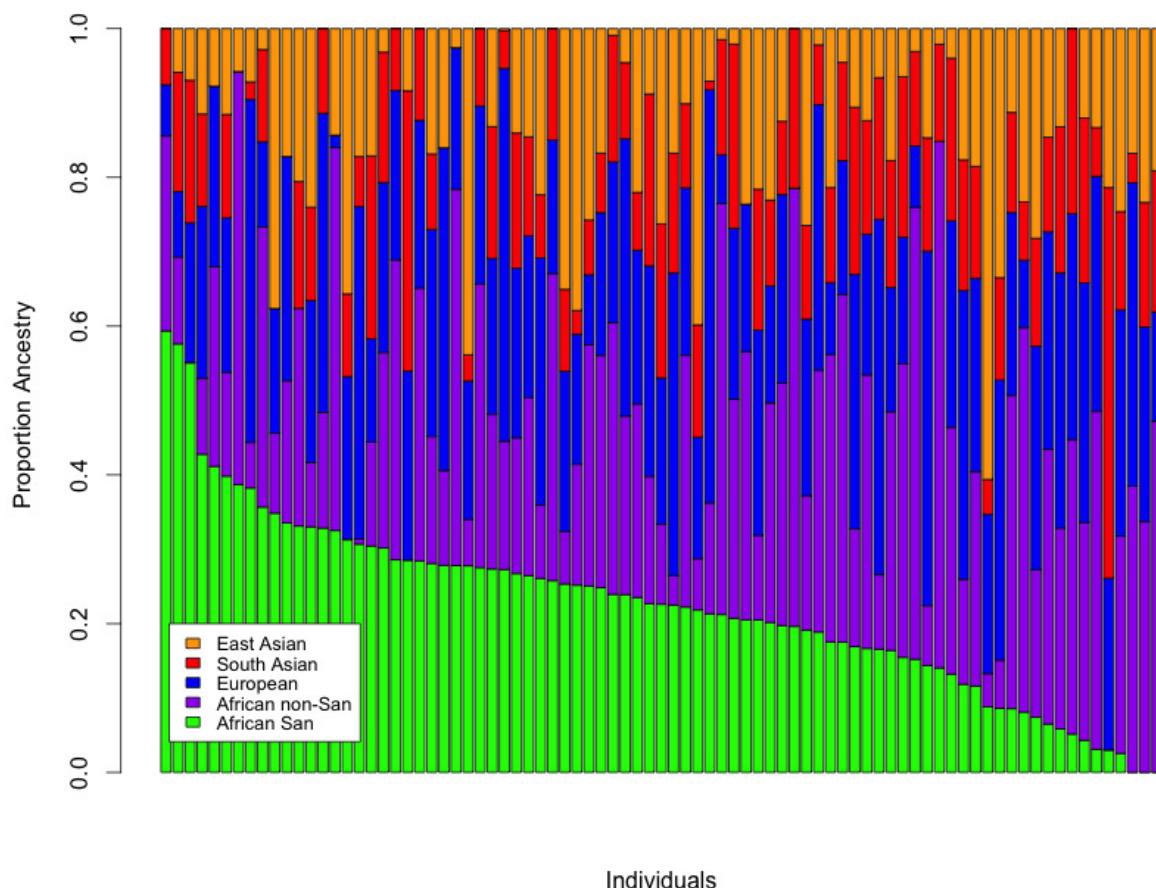


Figure 6.4: Proportions of ancestry, as determined by ADMIXTURE analyses of the 95 AIMs, observed in each of the SAC FES individuals that were examined.

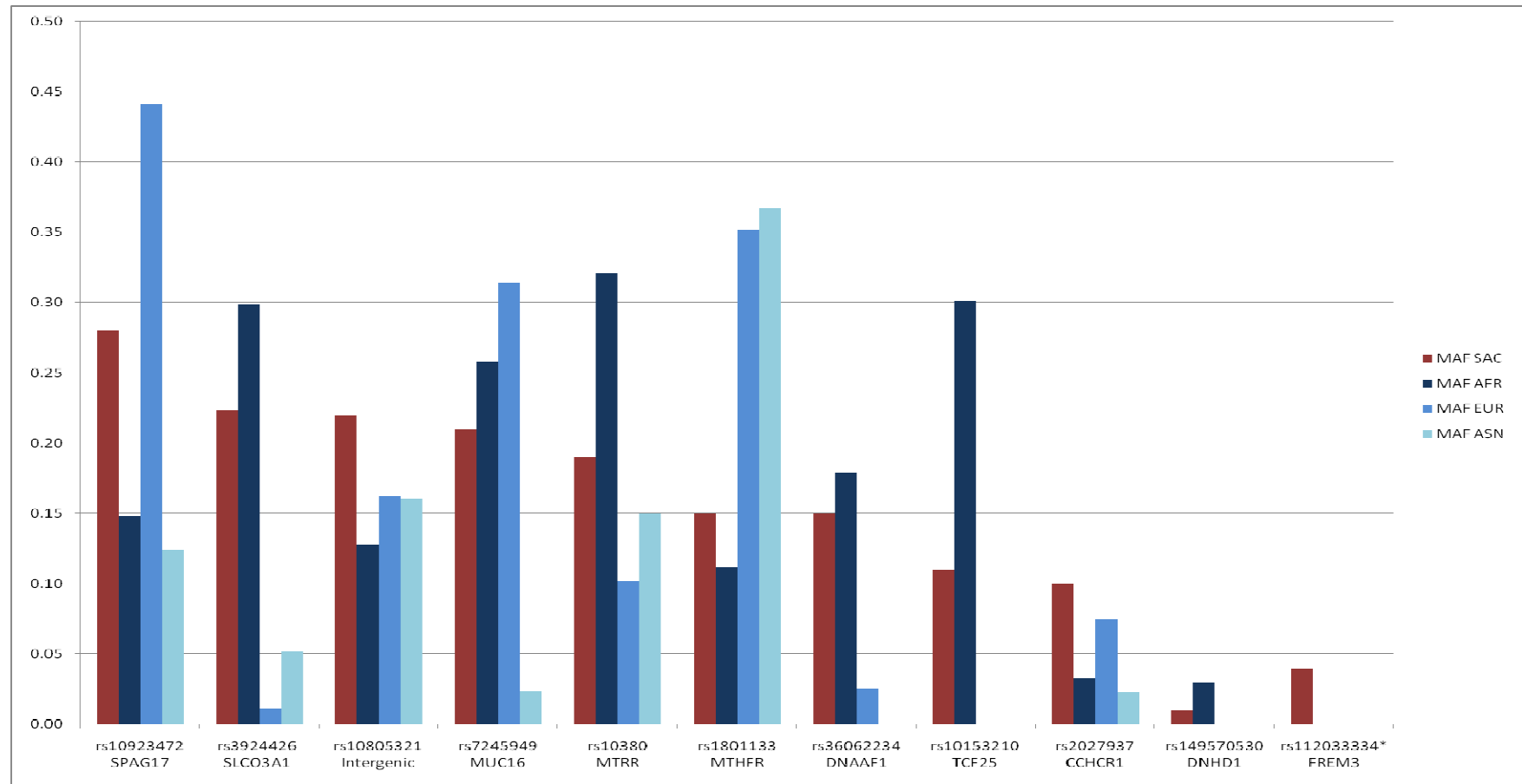


Figure 6.5: Minor allele frequency comparisons of the eleven polymorphisms that were significantly associated with antipsychotic pharmacogenomics phenotypes in the current study. Frequencies for the SAC FES population (n=82) and the African (n=492), European (n=758) and Asian (n=286) populations from the 1000 Genomes Project are shown. AFR: African, ASN: Asian, EUR: European, SAC: South African Coloured, MAF: Minor Allele Frequency, *There was no data available for this SNP on the 1000 Genomes Project browser.

6.5 Discussion

The current study utilised a unique strategy to identify the genetic variants involved in antipsychotic treatment response outcomes in a South African FES cohort. This approach made use of a well characterised cohort in combination with comprehensive genomic data. To our knowledge, this is the first study to utilise WGES for the identification of variants contributing to antipsychotic treatment outcomes. Although, the cost of WGES may be prohibitively expensive in the context of Africa, this study made use of a prioritisation strategy which encompassed the exome sequencing of only a subset of patients, thereby dramatically decreasing the associated costs. This strategy identified patients on extreme ends of the treatment response spectrum for exome sequencing in order to identify variants contributing to the phenotype of interest, as has been previously described [353,362–364]. In addition to this, sequencing of two siblings who were non-responders allowed for the subtraction of all variants that were not common to these two samples. This decreased the number of variants that needed to be considered in the downstream analyses. Furthermore, the genotyping strategy utilised in this study accounted for the fact that exome sequencing examines only the coding regions of the genome by including non-coding variants that have been associated with antipsychotic treatment response traits in previous GWAS. The panel of variants that was prioritised for genotyping in the entire cohort was also supplemented with variants from past candidate gene studies to make allowance for the fact that NGS may not be able to detect variation in all candidate antipsychotic genes (refer to Chapter 4). To our knowledge, this genotyping panel is the largest and most comprehensive that has been used to examine antipsychotic treatment response in the African context.

The use of the 384-plex BeadXpress Assay allowed for medium throughput genotyping, which provided the opportunity to examine a relatively large number of variants at an affordable price. Additionally, the size of this assay also facilitated the genotyping of a panel of AIMs. Analysis of these AIMs in the FES cohort revealed that the proportion of ancestry in the SAC individuals varied substantially (Table 6.4 and Figure 6.4). Thus, in order to eliminate the occurrence of false positives as a result of population stratification, genotyping of these AIMs in the FES cohort was essential [357]. The incorporation of the AIMs in the genotyping strategy permitted the inclusion of the SAC samples in our analyses. This is important as not only are the SAC the largest population group in the Western Cape (South Africa) where this study took place [365], but their admixed nature allows for the identification of variants from several different populations [242]. As shown in Figure 2.8, the allele frequencies of pharmacogenomic variants can differ substantially between populations. Therefore, certain variants may confer a risk for pharmacogenetic traits, but the risk that these variants confer may be specific to that population group. Examination of the highly admixed SAC population therefore allows for the unique opportunity to identify variants from several different population groups. This is highlighted in Figure 6.5, where the MAF of the majority of the significantly associated variants in the SAC occur at an intermediate frequency between the African, European and Asian populations. Although the variants identified may not have relevance to all populations (e.g. rs10153210), examination of the SAC allows for the identification of these variants without the need to sequence multiple populations.

The analysis of this unique cohort, in combination with the carefully constructed genotyping strategy, allowed for the identification of eleven variants that were significantly associated

with treatment response outcomes, even after correction for multiple testing of the 228 variants included in the association analyses. As mentioned in Chapter 2.3.3, only one antipsychotic GWAS has identified a variant that reached the genome-wide significance level required after adjustment for multiple testing and this study also utilised a FES cohort [203]. Thus, the results from the current study have served to highlight once again the importance of well characterised cohorts, particularly FES cohorts, for antipsychotic pharmacogenomic studies.

With reference to the specific associations that were detected, all but two of the variants were predicted to have a direct impact on the function of the gene in which they occurred [in the case of rs10153210, this was inferred from the fact that this variant tagged a missense variant, rs13338056 ($r^2=1$, LOD=3.33 in exome data)] (Table 6.3). The fact that these variants occur in protein coding sequences will allow for more direct applications in the future, as the specific effects of the identified proteins on antipsychotic treatment response can be examined. Although all of these variants were significantly associated with better treatment outcomes, with the exception of rs1801133 and rs10923472 which were associated with worse treatment outcomes, the observed effects that the variants had on the treatment outcomes varied substantially (Figure 6.3 and Appendix 10). The non-uniform effect that the variants had on treatment outcomes is to be expected as it has been reported that the treatment response trajectories observed between patients differ substantially, with patients showing different baseline PANSS scores and different rates of improvement [103]. The number of variants that were associated with treatment response outcomes combined with the observation that these variants acted in different manners to affect the treatment response trajectories, highlights once again the complexity of antipsychotic treatment response. These results suggest that many variants may act together to influence the wide array of treatment response phenotypes that are observed.

Focussing on the specific variants revealed that the two non-coding variants that were significantly associated with the changes in PANSS negative and general pathological scores (rs3924426 and rs10805321), although not directly impacting on the function of a protein product, have been associated with antipsychotic traits in past GWAS. The first variant, rs3924426, was previously associated with the occurrence of QT prolongation after treatment with iloperidone ($P=2 \times 10^{-6}$) [210] and the second variant, rs10805321, was associated with antipsychotic induced EPS ($P=9 \times 10^{-6}$) [212]. Although neither of these variants were associated with antipsychotic treatment response (which was not investigated by either of the studies), as shown in Table S2, antipsychotic candidate genes are often associated with both response to antipsychotic treatment and antipsychotic induced ADRs (e.g. *CYP2D6* and *DRD2*). The most likely reason for this is that the mechanisms involved in these antipsychotic treatment outcomes are interlinked [119]. This applies to the broader context where, as mentioned in Chapter 2.3.2, the genes involved in antipsychotic treatment outcomes overlap with the genes that have been implicated in the pathogenesis of schizophrenia [73]. Broadening the spectrum even further, it has recently been demonstrated that the genetics of psychiatric disorders are highly correlated [366]. Thus, the identification of variants involved in antipsychotic treatment response outcomes may have relevance to other aspects of antipsychotic treatment, as well as schizophrenia susceptibility or even susceptibility to other psychiatric disorders.

Examination of the significantly associated coding variants revealed that one of the variants, rs1801133 (A222V), has been reported to result in a 35% decrease in the enzyme activity of the protein encoded by an antipsychotic candidate gene, *MTHFR*, that is involved in the folate metabolism pathway [367]. Closer inspection of the remaining significantly associated variants (all identified by means of the VAAST analyses) revealed that another variant, rs10380 (H622Y) in *MTRR*, was also involved in the folate metabolism pathway [368]. Although both these genes have been implicated in schizophrenia susceptibility [368–370], no reports of an association with *MTRR* and antipsychotic treatment response were identified on the PubMed Database [371] (accessed 15th August 2013) [368,370]. Nonetheless, the folate metabolism pathway has been implicated in antipsychotic treatment response and rs1801133 has been associated with antipsychotic pharmacogenomic traits (Table S2) including a better response to antipsychotics [372,373]. Interestingly, our association with this variant was found in the opposite direction of the two other treatment response associations, both of which were performed in European descent individuals. As our cohort consists of predominantly SAC individuals, with a significant proportion of African ancestry (Figure 6.4), it is possible that the differences in the direction of effect may be attributed to differences in ancestry. This may be related to different patterns of LD and for this reason the surrounding variant landscape should be examined in the future, with particular reference to non-coding variants, which were not the focus of this study. Alternatively, due to the clinical heterogeneity of schizophrenia [374], the differences in ancestry may affect the way in which the schizophrenia and treatment response phenotypes are expressed. This has previously been reported as a potential explanation for the opposite effects that were observed with regards to the *COMT* rs737865 variant in the Xhosa population when compared to non-African populations [375]. This highlights once again the importance of examining the effects of genetic variants in different population groups, particularly in the context of Africa, which has been under-represented in terms of antipsychotic pharmacogenomic research [244].

With regards to the remaining seven significantly associated variants, six were associated with the changes in PANSS scores for the general pathological or negative symptom domain, which has been reported to show little improvement after antipsychotic treatment [15,83], while the remaining variant was associated with treatment refractoriness. To our knowledge, none of the genes whose functions were affected by these variants have previously been associated with antipsychotic treatment response traits or schizophrenia. However, one of the variants (rs112033334) was predicted to change the function of a gene (*FREM3*), which has been associated with major depressive disorder [370]. As mentioned above, the symptoms and genetics of psychiatric disorders are interlinked and more specifically, schizophrenia and major depressive disorder have been shown to overlap [47,366]. Therefore, this finding warrants further attention and highlights the fact the psychiatric genetic findings may have relevance to the broader spectrum of psychiatric disorders. Examination of the remaining six variants revealed that four of the variants (rs10923472 in *SPAG17*, rs149570530 in *DNHD1*, rs2027937 in *CCHCR1* and rs36062234 in *DNAAF1*) occurred in genes, which although poorly characterised with regards to their specific functions, have been implicated in microtubule associated functions [376–381]. This may be of importance as microtubules reportedly play a role in facilitating neuronal migration [382] and this process is vital to the development of a normal brain [383]. In addition, it has been reported that genetic variants which affect genes that are involved in

neuronal migration processes result in a wide array of diseases, including schizophrenia [384–386]. More specifically, variants in microtubule associated genes have also been implicated in the development of mental disorders [382,387,388]. This once again illuminates the likelihood that the genetics of antipsychotic treatment outcomes and psychiatric disorders overlap. Therefore, the specific functions of these genes, as well as the role that neuronal migration plays in both schizophrenia and treatment response phenotypes, should be investigated in the future. Of further significance, the variant associated with treatment refractoriness (rs2027938, A367T) occurred in a gene (*CCHCR1*) which is located in close proximity to the MHC region [389]. As mentioned in Chapter 2.1.3, the MHC region is one of the most well replicated findings in schizophrenia GWAS research [61] and has also been implicated in antipsychotic pharmacogenomic traits (Table S2), thus warranting further investigation. It should, however, be noted that this region is highly polymorphic and prone to false positives as a result of population stratification [390]. A further note of caution should be made with reference to the imprecise odds ratio (as can be seen by the wide confidence interval) observed for this variant, due to the relatively small sample size. This small sample size also has relevance to the associations observed with regards to rs112033334 and rs36062234, where there was only one homozygous individual for each of these variants that was driving the significant associations. These results should therefore be interpreted with caution and it will be necessary to genotype these variants in larger cohorts to confirm these findings.

With regards to the remaining two variants, one of the variants (rs7245949, T28911) was present in *MUC16*. It should be noted that it has been reported that *MUC16* is prone to false positives in WGES studies due to the large number of homologous sequences associated with the *MUC* genes [391]. Therefore, the significance of this variant should also be interpreted with caution. The remaining variant (rs10153210, which is in LD with the K53Q missense variant) was present in *TCF25*. This gene codes for a transcription factor expressed in the brain, which is involved in embryonic development [392]. As neuronal migration also occurs during development, the identification of variants involved in these processes may provide evidence for the neurodevelopmental hypothesis, whereby it has been postulated that abnormalities in brain development during gestation may contribute to the occurrence of the schizophrenia phenotype later in life [393]. Further validating these findings, the most recent schizophrenia exome sequencing study also identified a role for genes involved in neurodevelopment [69]. Of additional interest, both published PGC schizophrenia GWAS referred to in Chapter 2.1.3, identified significant associations with variants in the vicinity of *TCF4* (a neuronal transcription factor) [61,62], which may serve to provide further evidence for the role that transcription factors play in schizophrenia and antipsychotic response, as well as highlighting the importance of considering the effect of gene regulation and gene networks in antipsychotic treatment response outcomes. Taken together, the results from this study have highlighted the complexity of antipsychotic response, which appears to overlap with the genetics of schizophrenia. Nonetheless, this study has been successful in identifying novel and previously reported pathways were implicated in antipsychotic treatment response outcomes.

6.6 Conclusions

This study has demonstrated the utility of combining the use of well characterised, longitudinal cohorts and WGES technologies for the discovery of variants that may contribute to antipsychotic treatment response outcomes. Employing these data we were able to design an extensive panel of variants, including variants that have previously been associated with antipsychotic treatment response, for genotyping in larger cohorts of schizophrenia patients receiving antipsychotic medication. The use of this panel of variants in the FES cohort identified eleven variants that were significantly associated with antipsychotic treatment response outcomes. Nine of these variants were linked to a predicted change in function of a protein product (Table 6.3), thus providing direct information regarding the role that these variants may play in antipsychotic treatment response outcomes. With regards to the specific functional roles that these variants play, the folate metabolism pathway was implicated in the patterns of response that were observed with regards to the positive symptoms domain; while proteins associated with microtubules were implicated with regards to the negative or general pathological symptom domain. The identification of these microtubule associated genes, as well as *TCF25* which is reported to regulate genes involved in embryonic development, suggests a possible role for aberrant neuronal migration in relation to schizophrenia and response to antipsychotics. This highlights the validity of the neurodevelopmental hypothesis and brings to light the overlapping genetics of schizophrenia and antipsychotic treatment outcomes. It should also be noted that the majority of the significant associations were detected with respect to the negative symptom domain. This is of importance as antipsychotic treatments are less effective in treating this symptom domain [15,83]. Thus if the biological mechanisms involved in the lack of improvement in negative symptoms after antipsychotic treatment can be elucidated, novel strategies can be developed to optimise treatment protocols. In addition, this study revealed that the influence of genetic variants on treatment response outcomes varied. Thus, instead of a variant predicting whether a patient will respond or not respond to treatment, the genotype information is more likely to predict the treatment response trajectory of that patient. The number of variants that were significantly associated and the different affects that they have on treatment response highlight once again the complexity of antipsychotic response. If the results obtained from this study can be replicated in additional cohorts of schizophrenia patients receiving antipsychotic treatment, this information can be used to aid in our understanding of the mechanisms involved in antipsychotic treatment response outcomes and can ultimately be used to guide the optimisation of antipsychotic treatments.

Financial disclosure/acknowledgements:

The work reported here was supported by the following grants to the authors : BI Drögemöller: NRF research bursary and the L'Oréal-UNESCO for women in Science in Sub-Saharan Africa Fellowship; DJH Niehaus: South African MRC operating research grant; B Chiliza: South African MRC and Hamilton Naki Clinical Research Fellowship; L Asmal: None; GEB Wright: None; R Emsley: has participated in speakers/advisory boards and received honoraria from AstraZeneca, Bristol-Myers Squibb, Janssen, Lilly, Lundbeck, Organon, Pfizer, Servier, Otsuka and Wyeth. He has received research funding from Janssen, Lundbeck and AstraZeneca. This study was supported by a NEPAD grant, through the DST of South Africa; L Warnich: NRF operating research grants. This financial assistance is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily attributed to these funding sources.

Declaration by the candidate:

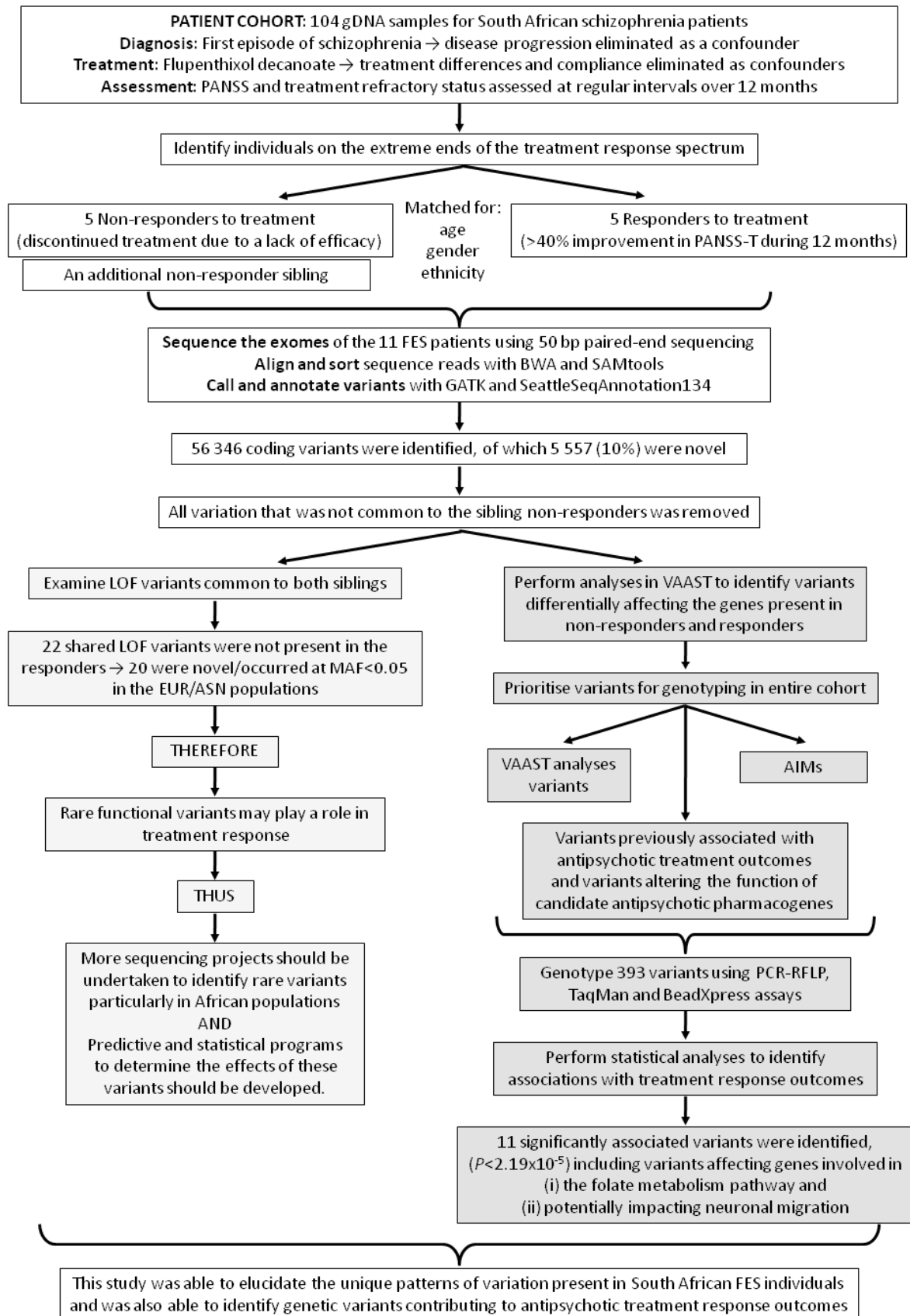
With regard to Chapters 6 (p 76-93), the nature and scope of my contribution were as follows:

Nature of contribution	Extent of contribution (%)
Funding and grant applications Design of project Identification of suitable service provider for genotyping gDNA quality control and preparation for shipping Gene and SNP based analyses for prioritisation of variants for genotyping Mining of GWAS data for prioritisation of variants for genotyping Identification of top antipsychotic pharmacogenes Mining of candidate antipsychotic pharmacogenes for prioritisation of variants Haplotype analyses of 1000 Genomes Project data Identification of candidate variants for genotyping Design of 384-plex BeadXpress genotyping assay and additional genotyping assays Sanger sequencing verification and genotyping of additional variants Critical assessment of genotype data Preparation of databases for statistical analyses Interpretation of data Writing of article	60%

The following co-authors have contributed to Chapter 6 (p 76-93):

Name	Email address	Nature of contribution	Extent of contribution (%)
Prof Louise Warnich	lw@sun.ac.za	Funding	10%
Dr Galen Wright	galen@sanbi.ac.za	Design of project Identification of suitable service provider Interpretation of data Critical assessment of article	
Prof Robin Emsley	rae@sun.ac.za	Funding	15%
Prof Dana Niehaus	djhn@sun.ac.za	Design of project Patient recruitment	
Dr Bonginkosi Chiliza	bonga@sun.ac.za	Patient diagnosis and treatment Clinical assessment of patients	
Dr Laila Asmal	laila@sun.ac.za	Critical assessment of article	
Prof Eileen Hoal	egvh@sun.ac.za	Design of AIMs	5%
Ms Michelle Daya	mdaya@sun.ac.za	ADMIXTURE analyses Critical assessment of article	
Prof Lize van der Merwe	lize.vandermerwe@mrc.ac.za	Statistical analyses Critical assessment of article	10%

Summary of Chapters 5 and 6



CHAPTER 7: Conclusions, study limitations and future directions

7.1 Conclusions

This study investigated a number of aspects regarding antipsychotic pharmacogenomics, which could broadly be divided into (i) the evaluation of the utility of WGES for antipsychotic pharmacogenomics in the African context and (ii) the subsequent utilisation of exome sequencing for antipsychotic pharmacogenomics in the South African setting. By critically examining the use of NGS in the context of Africa and subsequently applying these technologies, this study was able to successfully achieve the aims and objectives described in Chapter 2.5. The results of the first part of this study demonstrated the need for NGS technologies for antipsychotic pharmacogenomic research in the diverse populations of South Africa, while the second part of the study successfully implemented this genotyping strategy to identify eleven variants associated with antipsychotic treatment response, eight of which are novel.

7.1.1 PART 1: Evaluation of the utility of exome sequencing for antipsychotic pharmacogenomics in the African context

Mining of the scientific literature revealed that Africa has been under-represented in terms of genomic research. This applies particularly to psychiatric research, with no psychiatric GWAS reported in Africa at the time that the GWAS Integrator was searched [202]. Although the genomes of African individuals may not be well suited to SNP-chip analyses, the low LD and high levels of genetic variation make these individuals uniquely suited to WGES [250,262]. Due to the high rate of non-optimal treatment in Africa [296], the implementation of pharmacogenomics is urgently required. Unfortunately, due to the fact that populations residing within Africa have been under-represented in genomic research, little is known regarding the variation present in these individuals. This study has shown that NGS data from African individuals is more likely to uncover novel or uncharacterised variants, when compared to non-African populations. In addition, it has been shown that the frequencies of variants of relevance to pharmacogenomics are likely to differ substantially between population groups. Thus, these results indicate that WGES is required in order to elucidate the genetic factors contributing to antipsychotic response in the context of Africa.

However, due to the fact that it is known that important pharmacogenes such as the *HLA* and *CYP* genes are situated in complex regions of the genome, it was decided that the utility of NGS for variant detection in antipsychotic pharmacogenes should be critically examined. To do this, a subset of antipsychotic pharmacogenes and VIPs were analysed with regards to (i) sequence similarity to paralogues or pseudogenes, (ii) percentage GC content and (iii) percentage low complexity sequence. In addition each gene was examined to determine what percentage of the gene fell within the “inaccessible genome” as defined by the 1000 Genomes Project mask files. These analyses revealed that the inaccessibility of genes for NGS applications was highly correlated to the presence of paralogues/pseudogenes with high sequence similarity to the genes of interest ($P=8.02 \times 10^{-7}$), percentage GC content

($P=9.96 \times 10^{-5}$) and percentage low complexity sequence ($P=0.0002$). Furthermore, VIPs were more likely than antipsychotic pharmacogenes to be inaccessible ($P=0.035$), likely driven by the higher likelihood for VIPs to possess paralogues/pseudogenes ($P=0.029$). These results highlight the fact that intensively researched pharmacogenes, such as the *CYP* and *HLA* genes may not be ideally suited to NGS applications with current technologies. Therefore, alternate strategies may be required to examine these genes.

Taken together, these results have brought attention to the utility of WGES in Africa, while also providing information regarding the possible shortcomings associated with these technologies. With regards to the scarcity of genomic research in Africa, the future is looking bright with initiatives such as the H3Africa program lighting the way and playing a role in decreasing the disparities in genomic research [394]. Furthermore, the current lack of data pertaining to the diverse populations in Africa provides an opportunity for researchers to harness the ability of WGES to provide much needed information for these populations. However, while there is much to be gained from using these technologies, it is important to consider their shortcomings. Although it was demonstrated by this study that genomic complexities influence the ability of NGS technologies to call variants, these hurdles can in part be overcome by utilising information that has already been generated by past genetic association studies. In addition, the large number of databases that provide information regarding the genomic composition of the genes of interest (e.g. GC content, low complexity sequence and sequence similarity to other regions of the genome) [332,333,335,395,396], in combination with the information provided by the 1000 Genomes Project mask files [327], can serve as guidelines to indicate which areas of the genome may need to be examined with caution. By utilising information that has been generated by past studies together with novel information obtained from WGES studies, the missing pieces of the puzzle can slowly be added to the bigger picture of antipsychotic pharmacogenomics.

7.1.2 PART 2: The utilisation of exome sequencing for antipsychotic pharmacogenomics in the South African context

In order to identify the genetic variation contributing to antipsychotic treatment response in the South African context, exome sequencing was performed in eleven SAC FES individuals. To our knowledge, this study is the first to use WGES for pharmacogenomic and psychiatric applications in the context of Africa. To do this, an innovative strategy was utilised to overcome the prohibitive costs of WGES. This was done by making use of a subset of individuals on extreme ends of the treatment response spectrum, in combination with family data to reduce the number of variants that needed to be examined. Using this strategy allowed for a dramatic decrease in the associated costs and this approach may be used as a template for future genomic research that is performed in the resource limited settings of Africa. The use of this exome sequencing approach identified a total of 56 346 coding variants, of which 5 557 were novel (9.86%). The data generated from these analyses provided the first genomic information for South African SAC schizophrenia patients. Although genomic studies have been performed in the SAC population, these studies utilised SNP-chip based approaches [242,397], therefore limiting the ability of these studies to identify novel variants. Thus, this study is the first to use WGES in the SAC population to identify novel variants.

Examination of the global patterns of variation revealed that novel variants were unlikely to be shared between the non-responders and responders and 98% of the novel unshared variants were private and only present in one individual. Furthermore, LOF variants were more likely than other classes of variation to be novel. When examining only those LOF variants that were shared between the sibling non-responders and were not present in any of the responders, it was observed that there were 22 variants in 21 genes meeting these criteria. Of these 22 variants, 20 were novel or occurred at a $MAF < 0.05$ in the 1000 Genomes Project Asian and European populations, highlighting the complexity of antipsychotic response and the need to examine different population groups. In addition only one of these genes, *GRIN3B*, has a previous connection to antipsychotics. Although these analyses were not able to identify any variants that could completely explain non-response, they did highlight the importance of WGES in order to identify the variants contributing to non-response. These results also served to highlight the potential impact of rare variants on antipsychotic treatment response phenotypes and the need to develop methods to identify the role that these rare variants play. Due to the fact that rare variants are usually not shared between population groups [71], these results serve as yet another reminder that antipsychotic pharmacogenomic research needs to be performed in many different populations. This is particularly the case for the populations residing within Africa, as these populations have been under-represented in past genomic research and have high levels of rare variation [71].

Although, it appears that rare variation may play a role in antipsychotic treatment response outcomes, the effect of more common variants ($MAF > 0.01$) was also examined in this study. To do this, a panel of 393 variants was prioritised for genotyping in the entire cohort. This genotyping panel included (i) variants that have been associated with treatment response outcomes in past antipsychotic GWAS and candidate gene studies, (ii) variants present in the exome data that were predicted to alter the function of candidate antipsychotic pharmacogenes, (iii) variants in the exome data that were predicted to alter gene functions and occurred differentially in the non-responder and responder individuals and (iv) AIMs, to allow for the correction of population stratification in the statistical analyses. This comprehensive genotyping assay is, to our knowledge, the most extensive with regards to antipsychotic pharmacogenomics in Africa. This assay employed the use of the genomic data generated from the exome sequencing, while harnessing the information that has already been generated with regards to the genetic variants involved in antipsychotic treatment response. In so doing, this assay provided for the opportunity to identify novel genes involved in antipsychotic treatment response, while simultaneously allowing for the ability to replicate past associations.

With regards to the AIMs that were included in the genotyping panel, the information generated by these variants was integral to the results that were obtained in this study. Figure 6.4 clearly demonstrates that this group of individuals is highly admixed. Therefore, association analyses performed in these individuals are likely to be affected by population stratification and thus false positives may occur. However, as this population makes up a large percentage of the South Africa population [365], particularly in the Western Cape, where this study took place, it is not realistic to exclude this population group from genetic analyses, especially in cases such as this study where it is difficult to obtain large cohorts of FES patients on standardised treatment regimes. The AIMs used in this study proved

successful in allowing for the inclusion of the SAC samples in the association analyses. Furthermore, Figure 6.5 highlights that if the effects of population stratification can be accounted for, the admixed nature of the SAC may be of value for the identification of variants present in multiple populations, as described by Patterson *et al.* [242]. More specifically, the ability of admixed populations to detect pharmacogenetic variants which are of relevance to different population groups has previously been described [303,398]. These studies highlight the utility of admixed populations for pharmacogenetic applications, which was also demonstrated in the current study, where the SAC cohort provided the opportunity to identify variants that may be present in multiple South African populations thereby broadening the application of these variants in the context of the rainbow nation.

The statistical analyses performed in this cohort identified significant associations with eleven variants, nine of which were predicted to have an impact on the function of a protein product (Table 6.3). To our knowledge, eight of these functional variants have not previously been associated with antipsychotic treatment response outcomes and are therefore novel. The remaining two variants in non-coding regions have been associated with antipsychotic pharmacogenomic traits in past GWAS; therefore this study was able to once again bring to light the potential role that these variants play in antipsychotic outcomes. Examination of the specific variants that were significantly associated with antipsychotic treatment response outcomes revealed that different pathways were linked to different symptom domains, such that the folate metabolism pathway was implicated in improvement in positive symptoms, while genes potentially related to neuronal migration were associated with the negative or general pathological symptoms. This highlights that due to the non-uniform nature of schizophrenia [42], the different symptoms that are present in schizophrenia patients may need to be treated independently of one another. With regards to positive symptoms, due to the fact that evidence was provided that dysregulation of the folate metabolism pathway may influence treatment response, supplementation with folic acid may improve treatment regimes [370]. In contrast, the identification of the role that variants may play in influencing neurodevelopment processes can improve our understanding of negative or cognitive symptoms. As antipsychotic treatments are less effective for the treatment of these symptoms [15,83], this information may eventually be used in the development of drugs that target these symptoms. It should also be noted that the effect that the different variants were shown to have on treatment response trajectories varied substantially. These results reiterate once again the complexity of the genetics of antipsychotic treatment response.

Interestingly, both folate metabolism and neuronal migration have previously been implicated in schizophrenia susceptibility [69,368,393], which suggests that antipsychotic treatment response and schizophrenia are interlinked. Therefore, an improved understanding of antipsychotic response may improve our understanding of schizophrenia. Further evidence is provided for the similarities in the genetics of antipsychotic treatment response and schizophrenia by the fact that this study implicated both rare and common variants in treatment response outcomes. This is in accordance with the literature relating to the genetics of schizophrenia as past genomic studies examining schizophrenia have provided evidence for the role that both rare [65–68] and common variants [61,62] play in the development of the disorder. Due to the fact that both of these types of variation may be involved in antipsychotic treatment response and schizophrenia, different approaches

may be required to study the effects of these variants. In terms of rare variants, which are likely to be population specific [71], studies using sequencing technologies to examine several different populations may be required. However, common variants are likely to be shared across different population groups [399]. Furthermore, the contribution of common variants to schizophrenia susceptibility have been shown to overlap between African and European populations [400]. Therefore, the results obtained from the examination of the role that common variants play in treatment response in one population may allow for extrapolation onto other populations.

In closing, this study identified both novel and known genes that may be involved in antipsychotic treatment response. The identification of these genes should aid in improving our understanding of both the biological underpinnings of schizophrenia and antipsychotic treatment response outcomes. By using this information in future studies we may be able to move forward in the quest towards the use of genetic information to predict the treatment response outcomes of patients prior to treatment. This should play a role in optimising treatment regimes and reducing the socio-economic burdens placed on patients and their caregivers as a result of non-optimal antipsychotic treatment.

7.2 Study limitations

This study has a number of characteristics that makes it valuable to the field of antipsychotic pharmacogenomics. These strengths are related to (i) the unique and well characterised cohorts used in this study; (ii) the high throughput genomic technologies utilised and (iii) the application of these research strategies in the context of Africa. However, in conjunction with these strengths, there are also limitations that need to be considered. The main limitations of this study include (i) a relatively small sample size, including the lack of a replication cohort, and (ii) the inability to examine all variants, including structural and non-coding variants, in the entire cohort of patients.

With reference to the sample size, as described in Chapters 2.1.3 and 2.3.3, in order to detect associations with variants of small effect sizes, large cohorts are required. To obtain such cohorts it will be important to form international collaborations with structures similar to the PGC, which focus on antipsychotic pharmacogenomics. It should, however, be noted that it is very difficult to obtain large cohorts of well characterised schizophrenia patients for pharmacogenomic studies and in comparison to the majority of other antipsychotic pharmacogenetic studies, the FES cohort used in this study is large [189]. Nonetheless, even though this study made use of a well characterised discovery cohort for which extensive longitudinal data was available, it remains possible that some associations may have been missed. By using larger cohorts, additional variants that are significantly associated with antipsychotic treatment response may be detected. It should also be noted that although the FES cohort utilised in this study was well characterised in terms of pharmacogenomics and no other antipsychotics, mood stabilizers and psychostimulants were permitted; other psychotropic medications such as benzodiazepines, antidepressants or anticholinergics were used in parallel to the flupenthixol decanoate treatment when required. Thus, it will be important in the future to assess the effects of these multi-drug therapies on treatment outcomes.

In addition to the requirement for bigger sample sizes, the replication of results is also an important consideration. With reference to the results obtained from this study, we have already utilised our extensive genotyping panel to examine the genetic variation in a cohort of FES individuals obtained from the Zucker Hillside Hospital (New York, USA), for which clinical data is currently being collected. These individuals (n=89) are predominantly African American (69%) and thus will serve as an interesting comparison to the SAC individuals, as both these populations are highly admixed, with a large proportion of African ancestry (Figure 6.4, [260]). Mention should also be made to the fact that even though AIMs were included to correct for ancestry differences, the non-homogenous nature of these individuals with regards to ethnicity may complicate the analyses of these populations. Another aspect of the American FES patient cohort that will make for an interesting comparison is the fact that these patients are receiving SGAs. Therefore it will be of interest to determine if the results obtained from our study, which examined treatment response with regards to FGAs, can be extrapolated onto the treatment response outcomes observed after treatment with SGAs. The information generated from this replication cohort should aid in obtaining a more comprehensive understanding of the role that the associated genetic variants play in antipsychotic treatment response.

The genotyping strategy utilised in this study has many advantages, however, there are certain disadvantages that need to be considered. The three main shortcomings associated with exome sequencing are the inability to reliably detect (i) variants in complex regions of the genome (as described in Chapter 4), (ii) non-coding variants and (iii) CNVs [340]. To aid in overcoming the issues associated with the inability of NGS to detect variation in all regions of the genome, we included variants in candidate genes that have previously been associated with antipsychotic treatment response. Furthermore, a study by our group has undertaken to examine the variation in other known candidate antipsychotic pharmacogenes, such as *CYP2D6*, using already validated techniques including long range PCR. It should also be noted that sequencing technologies have already advanced during the course of this study and the sequence read lengths have increased [401], with technologies such as Pacific Biosciences' single molecule real-time sequencing technology able to obtain average read lengths of 3 000 bp [402]. Thus the ability of NGS to identify variants in complex regions of the genome is also likely to have improved. With reference to the inability of exome sequencing to detect non-coding variants, at present the most cost effective strategy is to utilise information on non-coding variants from past GWAS. In this study, to account for the effects of non-coding variants, the most significantly associated variants from past antipsychotic GWAS were included in the genotyping panel and a study by our group is currently examining the regulatory significance of these variants using the ENCODE data, as described by Schaub *et al.* [403]. Lastly, with regards to CNVs, although it is possible to examine CNVs with the use of NGS data, the capture and sequence biases introduced by exome sequencing complicate these analyses [340]. Moreover, the small number of exomes sequenced by this study make these strategies even more unreliable. The fact that only a subset of exomes were sequenced also prohibited the detection of rare variation in all individuals. As rare variation may play an important role in antipsychotic treatment response (refer to Chapter 5), these variants should be investigated in the future.

Lastly, it is important to refer to the ethical implications of WGES, which may be particularly important in the context of Africa where guidelines are limited [404]. As mentioned in the

introduction, the use of WGES is often accompanied by incidental findings, which introduces ethical concerns. As the stigma associated with pharmacogenomic findings is likely to be less than the stigma associated with disease-related findings and pharmacogenomic findings are also more likely to be actionable [30], the return of such findings to study participants should be considered. It will, however, be important that such findings are validated both in terms of the presence of the variants and their effects on treatment outcomes. Although this study did not identify any variants that would currently meet these criteria, in the future it will be important address these issues, as detailed by Wright *et al* [404].

7.3 Future directions

This study has generated a number of interesting results that have opened up several future research avenues. These avenues of research include (i) obtaining larger cohorts of schizophrenia patients, (ii) the utilisation of additional genotyping strategies, (iii) the examination of additional pharmacogenetic traits, (iv) the inclusion of information relating to expression and epigenetic profiles and (v) the incorporation of pharmacogenomic information into the clinical setting.

In terms of the first avenue of research, large cohorts are a valuable commodity to obtain statistical power. In order to obtain these large cohorts it will be important to form consortiums, such as the PGC, which examine not only schizophrenia but also the treatment thereof. It should, however, be noted that these cohorts must be well characterised. Thus, as antipsychotic pharmacogenomic research moves forward, it is essential that appropriate study designs are implemented to ensure the success of large studies. One such approach may include the implementation of the NIMH's Research Domain Criteria [47,405]. These criteria suggest that as the symptoms, treatments and genetics of patients with psychiatric disorders overlap substantially [47,366] (Figure 2.1), it may be more accurate to place individuals on a spectrum according to the symptoms that they display and to subsequently identify measurable endophenotypes that span multiple psychiatric disorders [405]. Identifying endophenotypes specific to antipsychotic response across multiple psychiatric disorders may aid in increasing samples sizes. By collecting large cohorts of patients, additional variants and pathways that are involved in antipsychotic treatment outcomes may be identified and the results obtained from the current study can be further substantiated.

Although the current study utilised a comprehensive genotyping strategy, this strategy still had limitations with regards to the ability to detect all types of variation. Thus, in the future, in order to investigate all types of variation, genome sequencing of all the patients in this cohort should be performed. Unfortunately, this may not currently be possible due to the costs associated with such analyses. Thus, future studies may instead combine the information obtained from exome sequencing with genotyping strategies such as those employed by the Affymetrix Genome-Wide Human SNP Array [406], which contains approximately one million SNP and CNV probes that are spread throughout the genome. The generation of these data would allow for examination of both non-coding variants and CNVs, and may also be used to provide a more comprehensive picture of the complex ancestry contributions of the SAC population. In addition, due to the prohibitively expensive

nature of WGES, targeted resequencing strategies of prioritised regions may be utilised in order to identify rare variants.

The cohorts of patients and genotyping strategies described above should also be used to examine other aspects of antipsychotic treatment. As described in Chapter 2.2.3, the occurrence of antipsychotic induced ADRs is a serious concern and can result in an increase in the socio-economic burdens placed on patients and their caregivers [15]. Therefore if optimal antipsychotic treatment is to be achieved, these pharmacogenetic traits also need to be considered. If the genetic factors that cause ADRs can be identified, pharmacogenetic strategies can be implemented to avoid the occurrence of these ADRs. In so doing, compliance to antipsychotic treatment can be improved [119]. Thus, by examining all aspects of antipsychotic treatment response, including ADRs, pharmacogenetic strategies can be designed to obtain optimal treatment response without the need for the lengthy trial and error procedures described in Chapter 2.2.1 [86,87].

In addition to these aspects of research which are focussed on examining the effects of genetic variants, strategies that utilise high throughput data generated from studies examining the differences in epigenetic regulation and gene expression levels should also be considered. As mentioned in Chapter 2.1.3, schizophrenia is a complex phenotype that is influenced by both the environment and genetics. Furthermore, previous research has shown that the use of antipsychotics affects gene expression levels and epigenetic regulation [407]. Therefore, these avenues of research are important and several NGS strategies exist that can facilitate this research. These include (i) RNA-seq, which examines the transcriptome and can thus be used to examine gene expression and ncRNA levels; (ii) ChIP-seq, which sequences the areas of the genome that are bound to proteins such as transcription factors, so as to identify transcription factor binding sites; (iii) Meth-seq, which can be used to identify the regions of the genome that are methylated and thus inaccessible to transcription factors; and (iv) DNase-seq, which can be used to assess the chromatin assembly and can thus be used to determine the accessibility of chromatin for transcription purposes [59,403,408].

The major hurdles associated with the implementation of these technologies are related to the fact that gene expression levels and epigenetic changes vary across cell types and are usually tissue specific [408]. Unfortunately, studies relating to the treatment of schizophrenia ideally require brain tissue, the use of which is prohibited by the inaccessibility of this organ [409]. Although the use of peripheral tissue, such as blood, may act as an indication of the gene levels in the brain, it is difficult to determine how accurate a representation this is [59]. Thus alternative methods such as the use of post-mortem brains, human induced pluripotent cells (hiPSC) and animal models may be more reliable. Unfortunately, although these methods have been applied to schizophrenia and antipsychotic research and interesting findings have been reported [407,409–411], all these methods have limitations. With regards to the post-mortem brain, not only is it difficult to obtain post-mortem brain samples, but the gene expression and epigenetic changes may be confounded by factors such as disease progression and treatment history [410]. Although the use of animal models does allow for the ability to manipulate genes and determine the effect of antipsychotic treatment on the animal, it is very difficult to model schizophrenia and to assess symptoms such as hallucinations and delusions in animals [59,412,413]. Thus,

it seems that the hiPSC approach may yield the most applicable results for antipsychotic pharmacogenomics. This innovative strategy can reprogram the fibroblast cells from schizophrenia patients into hiPSC, which can subsequently be differentiated into neurons [414]. Furthermore, it has been reported that the antipsychotic treatment of these hiPSC derived neurons can improve their neuronal connectivity [414]. Thus the use of hiPSC may improve our understanding of the mechanism of action of antipsychotic drugs, as well as to provide insight into why certain individuals do not respond to this treatment.

Together with the data generated by this study, this future research can be utilised to develop predictive tests, such as the one described in Table 1.1, that can be used to optimise antipsychotic treatment (Figure 2.7). It should, however, be noted that these predictive tests are likely to be more complicated than the one described in Table 1.1, due to the complexity of antipsychotic treatments. Instead of only providing recommendations regarding the dose of medication to use, these tests will also need to provide recommendations regarding the type of antipsychotic to use. Furthermore, due to the complexity of antipsychotic response, these tests may need to harness the power of sequencing technologies in order to obtain comprehensive pharmacogenomic information. Nonetheless, if these tests can be validated with regards to practicality, the treatment of schizophrenia can be improved and the associated burden and stigma can be reduced. As mentioned in Chapter 3, the use of pharmacogenetic tests can reduce the costs associated with ineffective treatments [287–291]. This is particularly important in the context of South Africa with its lack of resources, unique population groups and cultural considerations. In addition, if evidence can be provided regarding the biological underpinnings of schizophrenia and treatment response, the stigma can be reduced and individuals will be more likely to seek treatment. Together these strategies can be used to alleviate the unequal burdens of disease caused by schizophrenia in LMIC.

APPENDIX 1: Analyses performed in GALAXY

In order to obtain the data required for the analysis of the genomes/exomes examined by Schuster *et al* [280], the following was performed in GALAXY [306]:

- Click *Shared data/data libraries* > *Bushman* library (refer to Table S1 for a description of the individuals present in this dataset)
- Select *All SNPs in personal genomes*
- Click *Go*
- Return to homepage and select the *Bushman* library
- Click *Filter and sort* > *Filter*
- Select the following:
 - Filter: The bushman library dataset
 - C1='chr x ' and C2>=y and C3<=z; where x is the chromosome where the gene of interest is located and y and z are the chromosomal positions between which the gene is located. These positions were obtained for each of the genes from PharmGKB's gene boundary positions i.e. PharmGKB sets gene boundaries by expanding the mRNA boundaries by no less than 10 000 bases upstream (5') and 3 000 bases downstream (3') to allow for potential regulatory regions.
- Click *Execute*
- Click *Edit attributes* and rename with the “gene name”
- Click *Get data* > *UCSC Main*
- Select the following:
 - Assembly: Mar 2006 (NCBI36/hg18)
 - Group: Variation and repeats
 - Track: SNPs 130
 - Position: chr x :y-z (where once again where x is the chromosome where the gene of interest is located and y and z are the chromosomal positions)
 - Output format: BED
 - Send output to: Galaxy
- Click *Get output* > *Send query to Galaxy*
- Click *Edit attributes* and name the file with the “gene name_dbSNP130”
- Click *Operate in Genomic Interval* > *Subtract*
 - Then subtract “gene name_dbSNP” from “gene name” to get novel SNPs
- Click *Execute*
- Click *Edit attributes* and name “gene name novel SNPs”
- Click *Join, subtract, group* > *Join 2 datasets*
- Select the following:
 - Join “Gene name” using C2 with “Gene name_dbSNP” to get the rs numbers for the know SNPs
- Click *Execute*
- Paste list into Microsoft Excel (2007)
- Determine the effect of the identified variants using SIFT [308]

Table S 1: Thirteen ethnically diverse individuals utilised to compare pharmacogene variation

Individual identity	Continent of Descent	Geographic/Linguistic/ethnic group	Significance of individual	Date of sequencing		Sequenced region	
				2010	Prior to 2010	Genome	Exome
Desmond Tutu	Africa	Bantu (Xhosa/Tswana)	Archbishop and human rights activist	X		X	X
!Gubi	Africa	Tuu	Khoisan individual	X		X	X
G/aq'o	Africa	Juu (Ju/'hoansi)	Khoisan individual	X		X	X
D#kgao	Africa	Juu (Ju/'hoansi)	Khoisan individual	X			X
!Ai^	Africa	Juu (!Kung)	Khoisan individual	X			X
NA18507	Africa	Yoruba	HapMap Project participant		X	X	
NA 19240	Africa	Yoruba	HapMap Project participant		X	X	
Craig Venter	Europe	American	Involved in Human Genome Project		X	X	
James Watson	Europe	American	Co-discovered the DNA double helix structure		X	X	
NA 12891	Europe	American (Utah)	HapMap Project participant		X	X	
NA 12892	Europe	American (Utah)	HapMap Project participant		X	X	
Korean	Asia	Korean	First sequenced Korean individual		X	X	
Chinese	Asia	Chinese	First sequenced Chinese individual		X	X	

The different symbols within the names of the Khoisan individuals represent click consonants in the Khoisan languages [280].

APPENDIX 2: Next generation sequencing analyses of antipsychotic pharmacogenes

In order to obtain the data required for analysis of the antipsychotic pharmacogenes, the following was performed:

Commands performed in the Unix shell are indicated in grey font

- Identification of paralogues and pseudogenes using Ensembl BioMart [332]:
 - Select the following database and dataset: *Ensembl Genes 69* and *Homo sapiens genes (GTCh37.p8)*
 - Select the following filters: *Gene > ID list limit > HGNC ID(s)*
 - Paste in the names of the antipsychotic pharmacogenes
 - Select the following attributes: *Homologs > Paralogues > Human Paralog Ensembl Gene ID* and *% Identity with respect to query gene*.
- Identification of related pseudogenes and their corresponding sequences using NCBI's gene resource [333]:
 - Search for the relevant gene
 - Find the pseudogene information under *General gene information > Related pseudogenes*
 - Click on the pseudogene of interest
 - To obtain the sequence of the pseudogene click on *FASTA* under *Genomic regions, transcripts, and products*
 - Select and save the sequence
 - The gene sequences of interest were obtained in the same way.
- Identification of the percentage sequence similarity between genes and related pseudogenes using mVISTA [334]:
 - Give the number of pseudogenes that need to be aligned to the gene of interest under *Total number of sequences > Submit*
 - Upload and submit the sequences of the pseudogene(s) and relevant gene of interest.
- Calculation of GC content using Ensembl BioMart [332]:
 - Select the following database and dataset: *Ensembl Genes 69* and *Homo sapiens genes (GTCh37.p8)*
 - Select the following filters: *Gene > ID list limit > HGNC ID(s)*
 - Insert the names of the antipsychotic pharmacogenes
 - Select the following attributes: *Features > Gene > Ensembl Gene ID* and *% GC content*.

- Determination of the gene co-ordinates for the genes of interest using Ensembl BioMart [332]:
 - Select the following database and dataset: *Ensembl Genes 69* and *Homo sapiens Genes (GTCh37.p8)*
 - Select the following filters: *Gene > ID list limit > HGNC ID(s)*
 - Insert the names of the antipsychotic pharmacogenes
 - Select the following attributes: *Features > Gene > Ensembl Gene ID* and *Gene Start (bp)* and *Gene End (bp)*.
- Calculation of the percentage low complexity sequence present in the genes of interest using RepeatMasker [335]:
 - Under *Sequence selection* select
 - *Genome/Assembly: Human – Feb 2009 hg19*
 - *Range: Gene co-ordinates*
 - *Result type: Masked genomic sequence*
 - *Masking format: n*
 - Under *Filtering* select:
 - *Repeat classes: All*
 - Click *Submit Query*
 - Download the sequence
 - Determine the number of base pairs that were masked using the fgrep utility


```
$ fgrep -o n gene_name | wc -l
```
 - Calculate the percentage low complexity sequence by dividing the number of masked base pairs by the total number of base pairs in the gene.
- Calculation of the percentage “accessible genome”, as defined by the 1000 Genomes Project coverage data [281]:
 - Download the “strict mask” bed file [327]
 - Separate the masked sequence into different chromosomes (command for chromosome 1 shown here)


```
$ grep chr1 strict_masked_file.txt > chr1_strict_masked_file.txt
```
 - Determine the number of base pairs that were masked for each gene in Microsoft Excel (2007)
 - Calculate the percentage strict masked sequence by dividing the number of masked base pairs by the total number of base pairs in the gene.

Table S 2: The percentage sequence similarity, GC content, low complexity sequence and “inaccessible genome” associated with each of the pharmacogenes

Antipsychotic Pharmacogenes						
Gene	Associated Antipsychotic Pharmacogenetic Trait	Reference	Paralogue/Pseudogene	GC Content (%)	Low Complexity Sequence (%)	“Inaccessible Genome” (%)
<i>ABAT</i>	Movement disorder	[16,171,176,183]		47.35	57.35	36.26
<i>ACACA</i>	Weight gain	[184]	> 70% sequence similarity	41.27	43.88	29.66
<i>ADIPOQ</i>	Weight gain	[178,180,181]		43.88	41.43	24.48
<i>ADRA1A</i>	Treatment response, Movement disorder, Weight gain	[16,171,176,178,181]		41.26	40.72	13.48
<i>ADRA2A</i>	Movement disorder, Weight gain	[16,121,127,171,176,178–181]		59.81	4.63	48.11
<i>ADRB3</i>	Weight gain	[127,178,180,181]		59.22	10.11	41.16
<i>AJAP1</i>	Treatment response	[73]		52.75	28.54	15.50
<i>AKT1</i>	Treatment response	[16,176]	> 70% sequence similarity	64.08	5.86	22.47
<i>ALDH9A1</i>	Movement disorder	[16,171,176,183]		40.34	57.86	33.60
<i>ANKK1</i>	Treatment response, Movement disorder, Hyperprolactinemia	[121,187,193,194]		50.93	39.67	9.12
<i>ANKS1B</i>	Treatment response	[16,73,100,176,177]		37.84	55.86	21.26
<i>APOA4</i>	Weight gain	[181]		57.68	15.71	0.00
<i>APOA5</i>	Weight gain	[16]		60.22	0.00	1.31
<i>APOE</i>	Weight gain	[16,176,181]		61.22	27.18	48.97
<i>APOL3</i>	Weight gain	[16]		47.22	43.98	23.18
<i>ATP2B2</i>	Treatment response	[73]	> 70% sequence similarity	49.60	42.87	11.21
<i>BAG3</i>	Treatment response	[73]		49.77	26.25	23.17
<i>BDNF</i>	Treatment response, Movement disorder, Weight gain	[16,73,127,171,176,179–181,183,195,199]		40.24	32.41	26.02
<i>CELFA</i>	QT prolongation	[188]		52.35	20.06	10.30
<i>CERKL</i>	QT prolongation	[16,176,188,415]		37.01	44.16	12.13

<i>CHAT</i>	Treatment response	[16]		47.82	45.34	14.53
<i>CHRNA7</i>	Treatment response	[100]	> 70% sequence similarity	41.69	60.28	34.20
<i>CHST8</i>	Treatment response	[16,176,177,196]		50.64	43.90	24.33
<i>CNR1</i>	Treatment response, Weight gain	[16,127,176,180,181]		39.10	17.78	9.04
<i>CNTF</i>	Treatment response	[150,188]		41.63	7.50	0.00
<i>CNTN4</i>	Treatment response	[187]		37.97	40.71	17.25
<i>CNTNAP5</i>	Treatment response	[16,73,100,176,177]		38.93	45.34	17.17
<i>CYBA</i>	Agranulocytosis	[176]		61.89	28.39	49.31
<i>CYP17A1</i>	Movement disorder	[176,185]		52.48	19.27	16.35
<i>DECR1</i>	Movement disorder, QT prolongation	[16,121]		38.15	52.45	29.17
<i>DISC1</i>	Treatment response	[73]		42.18	46.87	22.55
<i>DRD1</i>	Treatment response	[176,193]		47.95	2.01	6.42
<i>DRD3</i>	Treatment response, Movement disorder	[15,16,73,100,100,119–121,148,171,176,177,179,183–187,189–191,193,195]		41.05	45.23	20.91
<i>DRD4</i>	Treatment response, Movement disorder, Weight gain	[16,73,119,171,176,177,179–181,183,185,190,191,193,195]		67.02	21.56	65.10
<i>DTNBP1</i>	Treatment response	[73,100,176,198]		39.93	55.11	21.15
<i>EHF</i>	Treatment response	[16,100,176,177,196]		43.81	19.44	7.45
<i>EN1</i>	Treatment response	[100]		58.21	8.44	39.38
<i>EPHA6</i>	Treatment response	[73]		35.41	47.94	18.48
<i>ESR1</i>	Movement disorder	[183,185]		39.66	44.55	22.32
<i>FAAH</i>	Weight gain	[180,181]	> 70% sequence similarity	55.38	31.74	30.19
<i>GABRA3</i>	Movement disorder	[16,171,176,183]		37.38	69.10	29.82
<i>GABRA4</i>	Movement disorder	[16,171,176,183]	> 70% sequence similarity	35.28	29.95	5.99
<i>GABRB2</i>	Movement disorder	[16,171,176,183,185]	> 70% sequence similarity	35.96	38.87	4.82
<i>GABRG3</i>	Movement disorder	[16,171,176,183,185]	> 70% sequence similarity	40.79	49.83	14.87

<i>GFRA2</i>	Treatment response	[15,73,100,150,176,177,188,415]		52.15	37.04	42.36
<i>GHRL</i>	Weight gain	[178]		53.50	18.01	0.77
<i>GLI2</i>	Movement disorder	[183]		52.42	29.05	10.91
<i>GNB3</i>	Treatment response, Weight gain	[16,100,119,127,148,171,176,178–181,195]	> 70% sequence similarity	54.70	27.17	21.06
<i>GPHN</i>	Movement disorder	[16,171,176,183]		36.43	65.29	29.58
<i>GPR137B</i>	Treatment response	[16,176,177,196]		43.62	48.32	35.46
<i>GRIA4</i>	Treatment response	[15,73,100,150,177,188,415]		35.32	35.14	13.33
<i>GRM3</i>	Treatment response, Weight gain	[16,176,189,198]		36.97	36.34	9.35
<i>GRM7</i>	Treatment response	[198]		37.80	43.53	13.71
<i>GRM8</i>	Treatment response	[100,196]	> 70% sequence similarity	37.51	45.20	21.12
<i>GSK3B</i>	Movement disorder	[183]	> 70% sequence similarity	37.15	55.06	25.64
<i>GSTM1</i>	Movement disorder	[16,171,183–185]	> 70% sequence similarity	46.35	55.35	96.88
<i>HLA-B</i>	Agranulocytosis	[121,136]	> 70% sequence similarity	59.00	0.00	97.01
<i>HLA-C</i>	Agranulocytosis	[121,136]	> 70% sequence similarity	59.28	0.00	99.26
<i>HLA-DQA1</i>	Agranulocytosis	[121,136]	> 70% sequence similarity	39.77	44.87	77.50
<i>HLA-DQB1</i>	Agranulocytosis	[16,121,136,176,186,189]	> 70% sequence similarity	47.03	13.30	95.00
<i>HLA-DQB3*</i>	Agranulocytosis	[136]	> 70% sequence similarity	48.72	0.00	72.15
<i>HLA-DRB1</i>	Agranulocytosis	[121,136]	> 70% sequence similarity	40.74	33.79	99.84
<i>HLA-DRB5</i>	Agranulocytosis	[121,136]	> 70% sequence similarity	42.84	36.39	100.00
<i>HOMER1</i>	Treatment response	[100]		37.23	47.30	23.37
<i>HRH2</i>	Treatment response	[120,176,195]		50.08	45.91	11.86
<i>HS3ST2</i>	Treatment response	[73]		43.49	51.66	22.60
<i>HSPA1A</i>	Agranulocytosis	[121,136]	> 70% sequence similarity	59.76	0.00	97.37
<i>HSPG2</i>	Movement disorder	[183]		56.09	38.41	26.10
<i>HTR1A</i>	Treatment response	[16,73,100,148,171,176,179,189,195]		58.78	5.39	3.11

<i>HTR2A</i>	Treatment response, Movement disorder, Weight gain	[15,16,73,88,100,119,120,120,121,127,148,150,171,176,177,179–181,183–187,189–191,195]		39.51	47.07	10.47
<i>HTR2C</i>	Treatment response, Movement disorder, Weight gain	[15,16,73,88,119–121,127,171,176–181,183–191,195]		35.03	50.91	24.03
<i>HTR3A</i>	Treatment response	[16]		50.68	33.94	16.24
<i>HTR3E</i>	Treatment response	[100]	> 70% sequence similarity	47.50	52.20	48.48
<i>HTR4</i>	Treatment response	[100]		38.87	56.24	18.87
<i>HTR6</i>	Treatment response, Weight gain	[16,73,119,127,176,177,179–181,195]		52.93	46.62	25.33
<i>HTR7</i>	Treatment response	[73,100,150,177,188,189,415]	> 70% sequence similarity	40.61	65.52	25.72
<i>IL1A</i>	Treatment response	[16,148,176,177,196]		39.97	32.70	11.50
<i>IL1RN</i>	Treatment response	[195]		46.04	43.53	14.03
<i>INSIG1</i>	Weight gain	[181]	> 70% sequence similarity	44.41	14.61	16.36
<i>INSIG2</i>	Weight gain	[16,176,180,181]	> 70% sequence similarity	37.96	16.00	13.21
<i>KCNH5</i>	Treatment response	[73]	> 70% sequence similarity	36.85	51.55	19.58
<i>LEP</i>	Weight gain	[16,121,127,171,176,178–181,184,186,189]		49.04	32.51	15.55
<i>LEPR</i>	Weight gain	[16,121,127,176,180,181,184]		37.79	55.36	26.34
<i>MC2R</i>	Treatment response	[176]		42.77	50.27	27.66
<i>MC4R</i>	Weight gain	[181]		45.06	0.00	0.00
<i>MEIS2</i>	Weight gain	[16,176]	> 70% sequence similarity	38.26	20.18	9.76
<i>MPO</i>	Agranulocytosis	[121,136,176]	> 70% sequence similarity	55.13	23.75	11.96
<i>MTNR1A</i>	Movement disorder	[183]		43.16	52.49	31.47
<i>NEFM</i>	Treatment response	[176,179]	> 70% sequence similarity	49.83	15.47	27.61
<i>NOS1AP</i>	QT prolongation	[16,176]		42.72	40.10	14.21
<i>NOS3</i>	Movement disorder	[171,176,183,185]		57.60	26.29	26.91
<i>NPAS3</i>	Treatment response	[15,16,73,100,150,176,177,188,415]		38.23	25.91	10.12
<i>NPYSR</i>	Weight gain	[181]		38.17	14.27	20.55
<i>NQO2</i>	Agranulocytosis	[121,176]		46.72	51.82	37.82

<i>NR3C2</i>	Treatment response	[73]		38.42	37.60	17.68
<i>NRG1</i>	Treatment response	[16,73,171,176,179]		37.90	37.65	16.42
<i>NRG3</i>	QT prolongation	[188]		37.51	43.36	14.04
<i>NUBPL</i>	QT prolongation	[16,176,188]		38.49	67.00	36.22
<i>NUDT9P1*</i>	Treatment response	[15,73,100,150,177,188,415]		38.84	0.00	0.00
<i>OPRM1</i>	Movement disorder	[183,185]		38.81	47.55	20.81
<i>OXT</i>	Treatment response	[176]	> 70% sequence similarity	71.94	0.00	88.18
<i>PAICS</i>	Treatment response	[73]		39.79	44.69	31.87
<i>PALLD</i>	QT prolongation	[188]		39.87	34.15	17.04
<i>PAM</i>	Weight gain	[178]		35.92	33.00	12.11
<i>PDE7B</i>	Treatment response	[73,100]		38.31	38.57	17.50
<i>PIP5K1B</i>	Treatment response	[73]		40.19	43.01	14.93
<i>PKHD1</i>	Weight gain	[178]		38.33	41.09	17.43
<i>PMCH</i>	Weight gain	[16,178,181]	> 70% sequence similarity	33.09	0.00	29.22
<i>PON1</i>	Weight gain	[181]		39.93	57.64	21.92
<i>PPARG</i>	Weight gain	[176,180,181]		38.71	38.36	17.58
<i>PRKAA1</i>	Weight gain	[181]	> 70% sequence similarity	37.74	49.88	24.88
<i>PRKCA</i>	Treatment response	[100]	> 70% sequence similarity	43.87	38.30	21.12
<i>PTGFRN</i>	Treatment response	[73]		44.19	44.98	18.50
<i>RGS2</i>	Movement disorder	[16,171,176,187]		40.83	0.96	8.10
<i>RGS4</i>	Treatment response	[16,73,171,176,177,179]		39.47	6.75	2.27
<i>RGS9</i>	Movement disorder	[176,183,185]		47.68	43.29	23.15
<i>RLN3</i>	Weight gain	[176]		50.95	53.52	34.52
<i>SCAP</i>	Weight gain	[181]		48.04	54.09	41.73
<i>SCARB1</i>	Weight gain	[181]	> 70% sequence similarity	51.49	58.38	30.31

<i>SCARB2</i>	Weight gain	[181]		41.24	40.78	18.30
<i>SLC1A6</i>	Treatment response	[100]		45.74	66.78	44.36
<i>SLC26A2</i>	Treatment response	[16,176,195]		41.19	58.85	25.33
<i>SLC26A9</i>	Treatment response	[16,177,189,196]		52.02	21.37	9.59
<i>SLC6A11</i>	Movement disorder	[16,171,176,183,185]		47.00	42.39	8.12
<i>SLC6A2</i>	Treatment response	[100]		48.47	38.06	12.53
<i>SLC6A3</i>	Treatment response	[16,73,100]		54.61	23.85	23.45
<i>SLC6A4</i>	Treatment response, Movement disorder, Weight gain	[16,73,100,119,120,127,171,176,178–181,194,195]		46.87	26.48	22.47
<i>SLCO3A1</i>	QT prolongation	[16,176,188]		45.12	37.67	12.88
<i>SNAP25</i>	Treatment response, Weight gain	[16,73,127,171,176,178–181,195]		40.89	31.15	8.85
<i>SOD2</i>	Movement disorder	[121,148,171,176,183,185,189]	> 70% sequence similarity	42.37	45.75	39.31
<i>SREBF1</i>	Weight gain	[181]		58.04	26.72	28.69
<i>ST6GAL2</i>	Treatment response	[73]		41.34	33.77	12.29
<i>TGFB1</i>	Weight gain	[181,184]		52.41	50.01	61.39
<i>TNF</i>	Treatment response, Weight gain, Agranulocytosis	[16,121,127,136,171,176,178–180]		52.82	8.63	0.00
<i>TNR</i>	Treatment response	[15,73,100,150,177,188,415]		43.87	33.72	9.22
<i>TRPM1</i>	Treatment response	[100,176]		43.98	52.22	21.95
<i>UNC5C</i>	Treatment response	[73]		37.41	32.94	13.75
<i>XKR4</i>	Treatment response	[15,16,73,100,150,176,177,188,415]		40.23	40.75	15.11
<i>ZBTB20</i>	Treatment response	[73]		36.04	34.33	10.86
<i>ZNF804A</i>	Treatment response	[100]		34.04	48.90	14.72
Gene	Associated Antipsychotic Pharmacogenetic Trait		Paralogue/Pseudogene	GC Content (%)	Low Complexity Sequence (%)	“Inaccessible Genome” (%)
<i>ABCB1</i>	Treatment response, Weight gain	[16,120,127,171,176,177,179–182]	> 70% sequence similarity	37.05	51.55	19.39

<i>ADRB2</i>	Weight gain	[178]		50.61	0.00	13.97
<i>COMT</i>	Treatment response, Movement disorder	[16,73,100,119–121,171,176,177,179,183,185,187,189,193,195–198]		53.41	50.18	39.19
<i>CYP1A2</i>	Treatment response, Movement disorder	[15,16,119,120,171,176,177,179,183–186]	> 70% sequence similarity	52.03	32.79	24.48
<i>CYP2C19</i>	Treatment response	[88,179]	> 70% sequence similarity	38.88	80.16	62.63
<i>CYP2D6</i>	Treatment response, Movement disorder, Weight gain, QT prolongation	[16,88,119–121,127,171,176,177,179–181,183–186,188–192]	> 70% sequence similarity	62.68	0.00	100.00
<i>CYP3A4</i>	Treatment response	[120,176]	> 70% sequence similarity	39.62	39.33	46.57
<i>CYP3A5</i>	Treatment response	[120]	> 70% sequence similarity	40.47	48.48	31.64
<i>DRD2</i>	Treatment response, Movement disorder, Weight gain, Hyperprolactinemia, Neuroleptic malignant syndrome	[16,73,88,100,119–121,127,148,171,176–181,183–187,189–191,193–196]		48.36	33.70	9.33
<i>GSTP1</i>	Movement disorder	[16,183]	> 70% sequence similarity	63.05	2.84	25.74
<i>MTHFR</i>	Treatment response, Weight gain	[16,176,177,179–181,195]		54.53	27.37	22.32
<i>NQO1</i>	Movement disorder, Agranulocytosis	[16,136,171,176,183,185]		47.09	54.04	47.77

Other Very Important Pharmacogenes

Gene	Associated Antipsychotic Pharmacogenetic Trait	Reference	Paralogue/Pseudogene	GC Content (%)	Low Complexity Sequence (%)	"Inaccessible Genome" (%)
<i>ACE</i>	NA	NA	> 70% sequence similarity	54.48	33.20	23.52
<i>ADH1A</i>	NA	NA		35.55	23.47	30.08
<i>ADH1B</i>	NA	NA	> 70% sequence similarity	35.81	24.52	35.49
<i>ADH1C</i>	NA	NA		35.59	29.86	36.09
<i>ADRB1</i>	NA	NA		57.51	6.99	51.80
<i>AHR</i>	NA	NA	> 70% sequence similarity	34.56	25.26	9.64
<i>ALDH1A1</i>	NA	NA	> 70% sequence similarity	36.05	41.13	13.25
<i>ALOX5</i>	NA	NA		47.72	50.96	29.26
<i>BRCA1</i>	NA	NA	> 70% sequence similarity	44.09	57.07	58.64
<i>CYP2A6</i>	NA	NA	> 70% sequence similarity	53.42	20.61	98.57

<i>CYP2B6</i>	NA	NA	> 70% sequence similarity	44.63	60.12	73.13
<i>CYP2C9</i>	NA	NA	> 70% sequence similarity	37.78	71.96	56.82
<i>CYP2E1</i>	NA	NA		45.84	48.72	24.79
<i>CYP2J2</i>	NA	NA		39.79	54.75	11.37
<i>DPYD</i>	NA	NA		35.09	39.65	13.67
<i>F5</i>	NA	NA		38.10	41.82	12.60
<i>G6PD</i>	NA	NA		56.32	35.05	37.56
<i>GSTT1</i>	NA	NA	> 70% sequence similarity	52.22	44.13	65.34
<i>HMGCR</i>	NA	NA		37.88	26.14	19.06
<i>KCNH2</i>	NA	NA		59.57	19.36	19.09
<i>KCNJ11</i>	NA	NA		63.78	12.90	37.70
<i>NR1I2</i>	NA	NA		46.33	42.32	19.30
<i>P2RY1</i>	NA	NA		50.23	0.00	14.30
<i>P2RY12</i>	NA	NA		35.77	21.96	7.93
<i>PTGIS</i>	NA	NA		46.04	70.42	26.90
<i>PTGS2</i>	NA	NA		34.87	8.88	4.06
<i>SCN5A</i>	NA	NA		50.93	30.97	10.19
<i>SLC19A1</i>	NA	NA		59.34	24.42	41.20
<i>SLCO1B1</i>	NA	NA	> 70% sequence similarity	35.22	70.87	26.17
<i>SULT1A1</i>	NA	NA	> 70% sequence similarity	52.72	44.23	89.67
<i>TPMT</i>	NA	NA	> 70% sequence similarity	41.46	48.78	30.30
<i>TYMS</i>	NA	NA	> 70% sequence similarity	45.77	49.32	52.92
<i>UGT1A1</i>	NA	NA	> 70% sequence similarity	42.47	40.53	17.39
<i>VDR</i>	NA	NA		47.72	37.33	18.38
<i>VKORC1</i>	NA	NA	> 70% sequence similarity	55.01	41.11	53.17

*This is a pseudogene; Those genes that are shaded in grey highlight the genes whose sequences are more than 50% inaccessible as determined by the 1000 Genomes Project strict masking.

APPENDIX 3: Specified protocols, reagents and solutions

Specified protocols

Purification of DNA from whole blood using the QIAamp® DNA Blood Maxi kit (QIAGEN, Germany)

1. Pipette 500 µl QIAGEN Protease into the bottom of a 50 ml centrifuge tube
2. Add 5-10 ml of blood, mix briefly
3. Bring volume to 10 ml with PBS
4. Add 12 ml Buffer AL, invert 15 times, shake for at least 1 min
5. Incubate at 70°C for 10 min
6. Add 10 ml 100% ethanol, invert 10 times, shake
7. Transfer half of the solution onto the QIAamp Maxi column, which is placed in a 50 ml centrifuge tube
8. Centrifuge at 3 000 rpm for 3 min
9. Remove the QIAamp Maxi column, discard the filtrate, place the QIAamp Maxi column back in the 50 ml centrifuge tube
10. Load the remainder of the solution onto the QIAamp Maxi column
11. Centrifuge at 3 000 rpm for 3 min
12. Remove the QIAamp Maxi column, discard the filtrate, place the QIAamp Maxi column back in the 50 ml centrifuge tube
13. Add 5 ml Buffer AW1 to the QIAamp Maxi column
14. Centrifuge at 5 000 rpm for 1 min
15. Add 5 ml Buffer AW2 to the QIAamp Maxi column
16. Centrifuge at 5 000 rpm for 15 min
17. Place the QIAamp Maxi column in a clean 50 ml centrifuge tube, discard the collection tube containing the filtrate
18. Pipette 1 ml Buffer AE directly onto the membrane of the QIAamp Maxi column
19. Incubate at room temperature for 5 min
20. Centrifuge at 5 000 rpm for 2 min

Miller *et al.* [354] gDNA Extraction Protocol

1. Shake the tube with blood well to mix the contents and transfer the contents (± 10 ml) to a marked 50 ml polypropylene tube
2. Add ± 30 ml cold Lysis Buffer and mix by inversion
3. Place the tube on ice for 15-30 min and mix by inversion every 5 min
4. Centrifuge the tubes for 10 min at 1 500 x g (4°C)
5. Carefully discard the supernatant and keep the pellet. Pat slightly dry on paper
6. Add 10 ml cold PBS to the pellet, mix and centrifuge again for 10 min at 1 500 x g (4°C)
7. Carefully discard the supernatant and keep the pellet. Pat slightly dry on paper
8. Dissolve pellet in: 3 ml Nuclear Lysis Buffer, 50 µl Proteinase K (10 mg/ml), 300 µl 10% SDS
9. Shake very well and incubate overnight in a water bath at 56°C
10. Add 1 ml 6M NaCl to each tube and shake continuously for 1 min
11. Centrifuge for 20 min at 2 500 x g at room temperature
12. Transfer supernatant to a Falcon tube. Be careful not to transfer any of the pellet or foam. The supernatant must be clear
13. Add 3 volumes ice cold (-20°C) 99.9% ethanol to the supernatant in the Falcon tube and mix very carefully
14. A DNA bundle should form. Carefully hook the bundle out with a needle and place it in an Eppendorf tube that contains 1 ml 70% ethanol
15. Centrifuge at 1 400 x g for 5 min at 4°C

16. Carefully discard the ethanol and allow the pellet to dry
17. Dissolve the pellet in 100-200 μ l TE, depending on the size of the pellet

SureClean Quick-Clean Protocol (Bioline, UK)

1. Add 1 x volume of Quick-Clean to nucleic acid solution, vortex thoroughly
2. Incubate at room temperature for 10 min
3. Centrifuge at maximum speed in a microfuge for 10 min, discard supernatant
4. Add 100 μ l of 70% Ethanol, vortex for 30 sec
5. Centrifuge at maximum speed for 10 min, remove supernatant
6. Air-dry to ensure complete removal of ethanol
7. Resuspend pellet in 10 μ l of water

Big Dye v3.1 Sequencing Chemistry (Applied Biosystems™, California, USA)

1. Add 21 ng of purified PCR product
2. Add 1.3 μ l of Big Dye reaction mix
3. Add 2.7 μ l Half Dye mix
4. Add 3.3 μ mol of primer

Reagents and solutions

All solutions were brought to volume with distilled water.

10% Sodium Dodecyl Sulphate (SDS)

0.3468 M SDS (BDH Laboratory Supplies)

10X TBE Electrophoresis Buffer (pH 8.3) (All reagents supplied by Sigma-Aldrich (Pty) Ltd.)

0.0890 M Trizma® Base

0.0890 M Boric acid

0.0020 M Ethylenediaminetetraacetic acid

15% Dimethyl sulfoxide (DMSO)

15% DMSO (v/v) (Sigma-Aldrich (Pty) Ltd.)

70% ethanol

70% (v/v) ethanol (Sigma-Aldrich (Pty) Ltd.)

Agarose gels

1-3% (w/v) Agarose (SeaKem®)

0.5 µg/ml Ethidium Bromide (Sigma-Aldrich (Pty) Ltd.)

Cresol Loading Dye

2% (v/v) cresol stock solution (Sigma-Aldrich (Pty) Ltd.)

0.9933 M sucrose (Merck Chemicals (Pty) Ltd.)

Phosphate Buffered Saline (PBS) (pH 7.4) (All reagents supplied by Merck Chemicals (Pty) Ltd.)

0.0268 M KCl

0.1369 M NaCl

0.0080 M Na₂HPO₄

0.0015 M KH₂PO₄

APPENDIX 4: Exome data generation

Preparation of gDNA samples

The gDNA was extracted from whole blood samples with the QIAamp DNA Blood Maxi Kit (QIAGEN, Germany) according to manufacturer's instructions and eluted in 1 ml of AE Buffer. Thereafter the NanoDrop® ND-100 (Nanodrop Technologies Inc., Delaware, USA) was used to assess the quality, purity and concentration of each gDNA sample at 260 nm. Each sample was subsequently diluted with Qiagen Buffer EB (QIAGEN, Germany) to 50 ng/μl, of which 100 μl was aliquoted into an eppendorf tube and frozen at -20°C. All frozen gDNA samples were shipped on dry ice by DHL couriers. An export permit for biological substances was obtained from the South African Department of Health and this was included in the shipment along with a hazardous goods declaration form.

Quality control of gDNA samples

Once the gDNA samples were shipped to the HudsonAlpha Genomics Service Laboratory, they were run on a 0.8% agarose gel in 1X TBE at 100 V for 1 hr 45 min to assess the quality of the samples. Samples were given lab codes, LW0001-LW0011, and the resultant gel for the assessment of gDNA integrity is shown below in Figure S1. All samples were deemed intact. The Qubit® Fluorometer (Life Technologies, California, USA) was subsequently utilised to accurately determine the concentration of each sample and the total ng of gDNA available (Table S3)

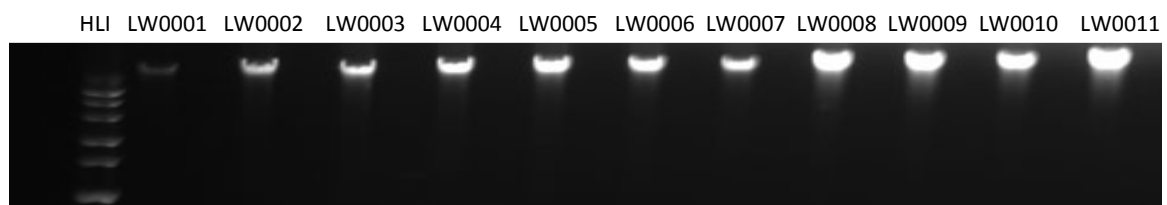


Figure S 1: The integrity of gDNA samples LW0001-LW0011

Table S 3: Quality assessments for the gDNA samples

Sample ID	Reported Conc (ng/μl)	Initial QC Qubit (ng/μl)	Calculated QC Qubit (ng/μl)	Reported volume (μl)	Remaining volume (μl)	Total (ng)
LW0001	50	16.4	49.2	150	140	6888.0
LW0002	50	16.3	48.9	150	140	6846.0
LW0003	50	16.6	49.8	150	140	6972.0
LW0004	50	16.5	49.5	150	140	6930.0
LW0005	50	16.8	50.4	150	140	7056.0
LW0006	50	14.4	43.2	150	140	6048.0
LW0007	50	13.5	40.5	150	140	5670.0
LW0008	50	19.9	59.7	150	140	8358.0
LW0009	50	16.2	48.6	150	140	6804.0
LW0010	50	16.7	50.1	150	140	7014.0
LW0011	50	21.7	65.1	150	140	9114.0

Conc: Concentration, QC: Quality control

Quality control of sequence reads

After the quality of the gDNA samples was verified, the samples were ready for processing. The samples underwent (i) sonification, end repair and subsequent quantification with a Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen™, California, USA); (ii) whole genome amplification and subsequent quantification, (iii) hybridisation with the Agilent SureSelect All Exon 50 Mb capture kit (Agilent Technologies, California, USA) and subsequent quantification with the Bioanalyzer (Agilent Technologies, California, USA) and (iv) 50 bp paired end sequencing on the HiSeq2000 (Illumina, California, USA).

The generated sequence reads were then aligned to the human genome reference sequence (hg19), using BWA [325] and the reads were sorted and indexed using SAMtools [347] (refer to Figure S2 for a graphic representation of bioinformatic pipeline used). The quality of the generated sequence reads was assessed using an in house quality control program at HudsonAlpha Genomics Service Laboratory [346]. This program is similar to the free software package, FastQC [416]. Information was provided for observed (i) fragment sizes, (ii) qualities, (iii) qualities per cycle and (iv) bases per cycle, for each individual (Figure S3). A summary of the sequence coverage and alignment is provided in Tables S4 and S5.

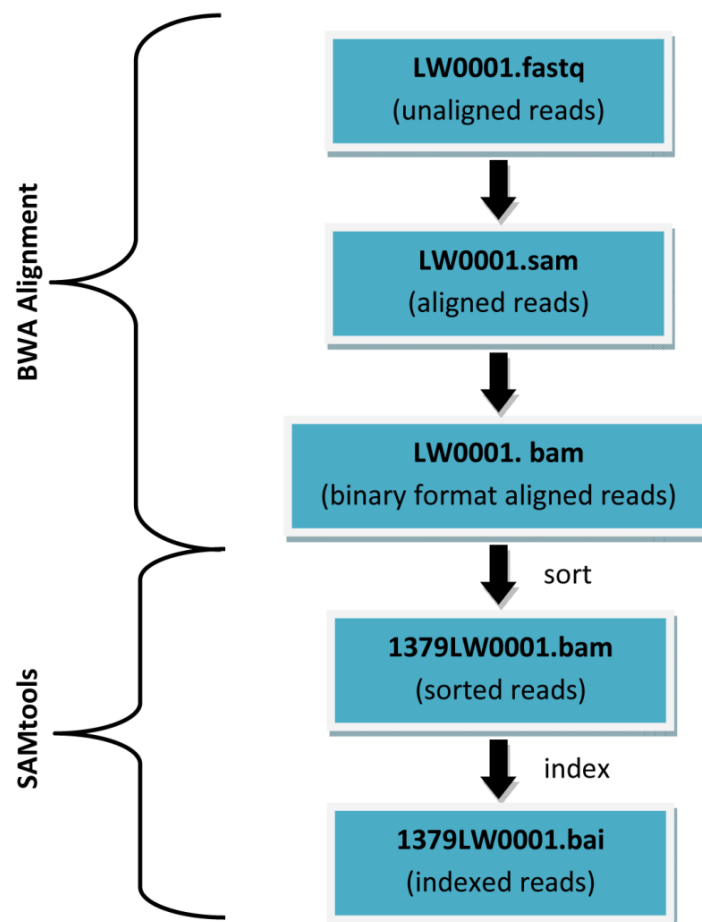


Figure S 2: Alignment and sorting of generated reads

Table S 4: Summary of sequence coverage obtained for the targeted exomes of the eleven FES samples

Sample	Median Read Depth	1x Coverage (%)	4x Coverage (%)	8x Coverage (%)	20x Coverage (%)	30x Coverage (%)
LW0001	72x	99.11%	97.41%	95.36%	89.53%	83.93%
LW0002	67x	98.94%	97.14%	94.88%	88.42%	82.00%
LW0003	82x	99.14%	97.81%	96.20%	91.53%	86.93%
LW0004	63x	99.09%	97.23%	94.87%	87.80%	80.63%
LW0005	81x	99.20%	97.69%	95.92%	90.96%	86.24%
LW0006	66x	99.21%	97.49%	95.31%	88.68%	81.99%
LW0007	82x	99.22%	97.78%	96.00%	90.79%	85.95%
LW0008	76x	99.55%	98.36%	96.58%	90.94%	85.83%
LW0009	67x	99.20%	97.64%	95.67%	89.55%	83.13%
LW0010	61x	98.91%	96.91%	94.45%	87.13%	79.63%
LW0011	63x	98.78%	96.85%	94.50%	87.70%	80.74%

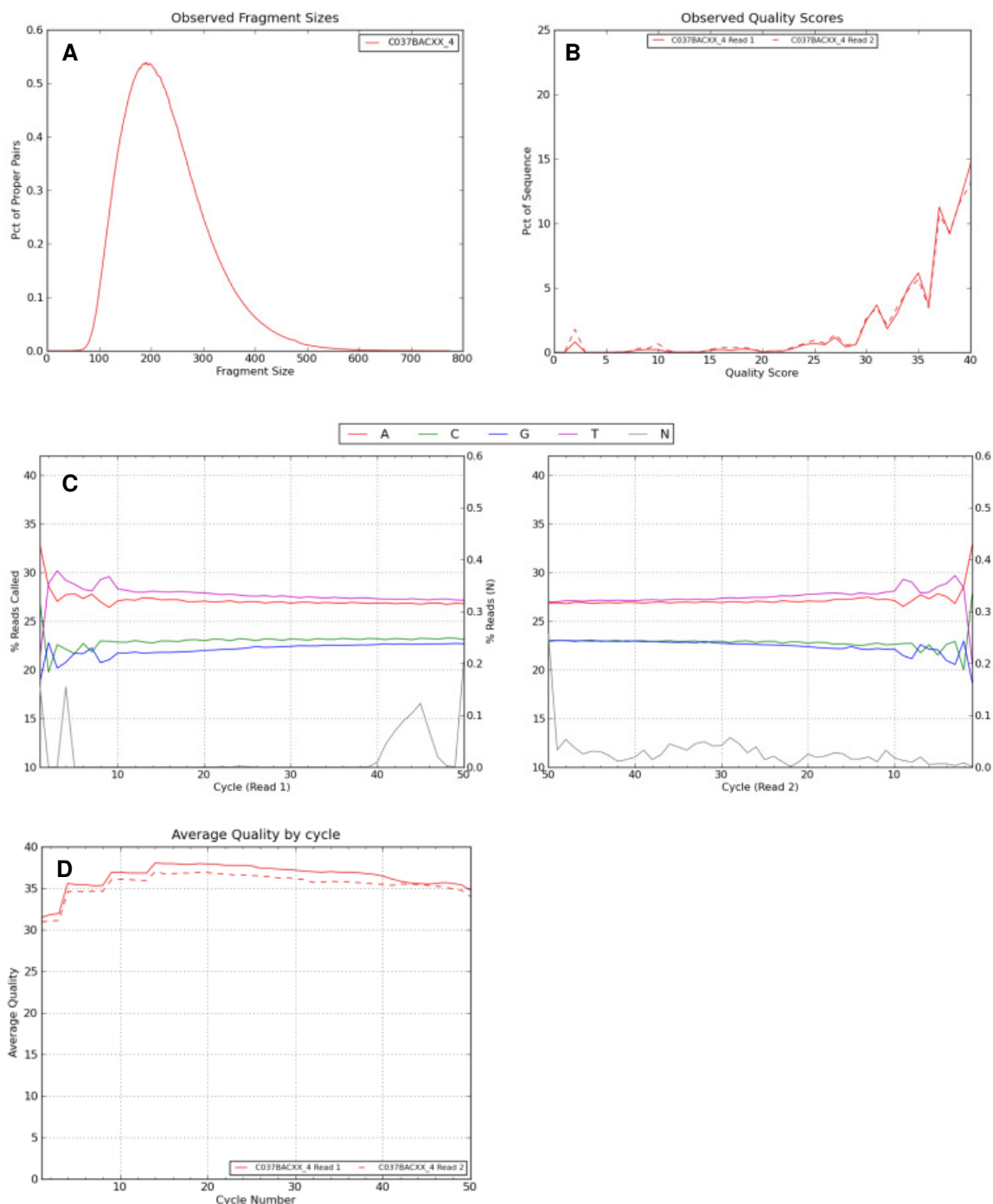


Figure S 3: Example of **A:** Observed fragment sizes; **B:** Observed quality scores; **C:** Observed bases per cycle; **D:** Observed quality scores per cycle.

Table S 5: Sequence reads generated and aligned to the hg19 reference sequence

Sample	# Fragments	Total sequence (bp)	Aligned pairs (proper)	Aligned sequence (%)	Track size (bp)	Sequence in track (%)	Sequence in track plus or minus 100bp (%)	Sequence in track plus or minus 200bp (%)
LW0001	113904678	11390467800	98.33%	99.00%	51646629	44.86%	61.60%	66.20%
LW0002	101842349	10184234900	98.45%	99.06%	51646629	46.39%	63.76%	68.47%
LW0003	125223080	12522308000	98.24%	98.99%	51646629	46.93%	65.22%	70.78%
LW0004	100075179	10007517900	98.12%	98.86%	51646629	44.95%	62.76%	68.35%
LW0005	118744308	11874430800	98.41%	99.05%	51646629	47.83%	66.64%	72.50%
LW0006	102500028	10250002800	98.18%	98.95%	51646629	45.48%	63.52%	69.35%
LW0007	119766615	11976661500	98.16%	98.96%	51646629	47.44%	66.50%	72.70%
LW0008	146696944	14669694400	97.97%	98.76%	51646629	32.81%	45.84%	50.04%
LW0009	104929467	10492946700	98.18%	98.88%	51646629	45.24%	62.92%	68.40%
LW0010	96405718	9640571800	98.23%	98.92%	51646629	44.47%	61.71%	66.90%
LW0011	106194950	10619495000	98.28%	98.97%	51646629	41.93%	58.22%	63.07%

APPENDIX 5: Variant calling and annotation of exome data

The exome analyses were performed using a variety of different programs. These specific analyses are documented below. In each case, the sample LW0001 is used to describe the analyses performed, however, these analyses were performed identically for all eleven FES exome samples.

Commands performed in the Unix shell are indicated in grey font

Transfer of data

- The data (bam and bai files) generated by HudsonAlpha Genomics Service Laboratory were downloaded using the internet file transfer program (i.e. ftp) onto a local laboratory computer.
- All data was backed in duplicate on two external hard drives, which were kept at separate locations.
- In addition, the data was transferred from the local laboratory computer to the “data” directory on the Rhasatsha High Performance Computer (HPC), at the Department of Engineering, Stellenbosch University, via the secure copy (scp) utility:

```
$ scp 1379LW0001.bam 14337185@head002.sun.ac.za:data
```
- To transfer data from the HPC onto the local laboratory computer the scp utility was also used:

```
$ scp 14337185@head002.sun.ac.za:file_name .
```
- Excluding the transfer of files to and from the HPC, all transferring of files to and from remote and local sites was performed using FileZilla [417].
- All software that was required for the downstream analyses was installed onto the HPC.

Submission of jobs to the high performance computer

- To submit jobs a submit file was made in vim:

```
$ vim file_name
```
- Each file contained the following header:

```
#!/bin/sh
#PBS -m abe
#PBS -M 14337185@sun.ac.za
```

Where abe instructs an e-mail to be sent when the job aborts, begins or exits.
- For each submit file the commands required for the job were entered and the file was saved:

```
$ ZZ
```
- The job was submitted to the queue on the HPC:

```
$ qsub file_name
```
- The status of the submitted jobs was checked:

```
$ qstat -r
```

The Genome Analysis Toolkit (GATK): variant calling

- To run the GATK [326] the required resources (Table S6) were obtained from: <ftp://gsapubftp-anonymous@ftp.broadinstitute.org> (Password: <blank>), in the folder bundle/1.2/b37 (updated 20/09/2011) and were downloaded into the folder /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/

Table S 6: The GATK resources

Resource	Description
human_g1k_v37.fasta and human_g1k_v37.fai	Human reference sequence
dbSNP_132.b37.vcf and dbSNP_132.b37.vcf.idx	The dbSNP release version 132
hapmap_3.3.b37.sites.vcf and hapmap_3.3.b37.sites.vcf.idx	HapMap genotypes
1000G_omni2.5.b37.sites.vcf and 1000G_omni2.5.b37.sites.vcf.idx	1000 Genomes genotypes

- A graphical representation of the GATK read recalibration and variant calling processes are shown in Figure S4. Submit files for the GATK analyses were created and submitted on the HPC using the following scripts:

Picard: MarkDuplicates

```
$ java -Xmx4g -jar /apps/picard-tools-1.50/MarkDuplicates.jar
INPUT=/export/home/14337185/data/1379LW0001.bam
OUTPUT=/export/home/14337185/LW0001_mkd.bam
METRICS_FILE=/export/home/14337185/LW0001_metrics.txt ASSUME_SORTED=true
CREATE_INDEX=true VALIDATION_STRINGENCY=LENIENT
```

The GATK: RealignerTargetCreator

```
$ java -Xmx1g -jar /apps/GenomeAnalysisTK-1.1-31-gdc8398e/GenomeAnalysisTK.jar \
-T RealignerTargetCreator \
-R /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/human_g1k_v37.fasta \
-o LW0001.intervals \
-l /export/home/14337185/LW0001_mkd.bam \
```

The GATK: TargetRealigner

```
$ java -Xmx4g -jar /apps/GenomeAnalysisTK-1.1-31-gdc8398e/GenomeAnalysisTK.jar \
-l /export/home/14337185/LW0001_mkd.bam \
-R /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/human_g1k_v37.fasta \
-T IndelRealigner \
-targetIntervals LW0001.intervals \
-o LW0001realignedBam.bam \
-compress 0 \
```

The GATK: CountCovariate

```
$ java -Xmx4g -jar /apps/GenomeAnalysisTK-1.1-31-gdc8398e/GenomeAnalysisTK.jar \
-R /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/human_g1k_v37.fasta \
-B:dbSNP.vcf /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/dbSNP_132.b37.vcf \
-l /export/home/14337185/LW0001realignedBam.bam \
-T CountCovariates \
-cov ReadGroupCovariate \
-cov QualityScoreCovariate \
-cov CycleCovariate \
-cov DinucCovariate \
-recalFile LW0001.recal_data.csv \
```

The GATK: TableRecalibration

```
$ java -Xmx4g -jar /apps/GenomeAnalysisTK-1.1-31-gdc8398e/GenomeAnalysisTK.jar \
  -R /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/human_g1k_v37.fasta \
  -l /export/home/14337185/LW0001realignedBam.bam \
  -T TableRecalibration \
  -o LW0001.recal.bam \
  -recalFile /export/home/14337185/LW0001.recal_data.csv \
```

The GATK: CountCovariate

```
$ java -Xmx4g -jar /apps/GenomeAnalysisTK-1.1-31-gdc8398e/GenomeAnalysisTK.jar \
  -R /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/human_g1k_v37.fasta \
  -B:dbsnp,vcf /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/dbsnp_132.b37.vcf \
  -l /export/home/14337185/LW0001.recal.bam \
  -T CountCovariates \
  -cov ReadGroupCovariate \
  -cov QualityScoreCovariate \
  -cov CycleCovariate \
  -cov DinucCovariate \
  -recalFile LW0001.recal2_data.csv \
```

The GATK:Rscript

```
$ java -Xmx4g -jar /apps/GenomeAnalysisTK-1.1-31-gdc8398e/AnalyzeCovariates.jar \
  -recalFile /export/home/14337185/LW0001.recal_data.csv \
  -Rscript /usr/bin/Rscript \
  -resources /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources \
  -outputDir /export/home/14337185 \
  -ignoreQ 5 \
```

The GATK: UnifiedGenotyper

```
$ java -jar /apps/GenomeAnalysisTK-1.1-31-gdc8398e/GenomeAnalysisTK.jar \
  -R /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/human_g1k_v37.fasta \
  -T UnifiedGenotyper \
  -l /export/home/14337185/LW0001.recal.bam \
  -l /export/home/14337185/LW0002.recal.bam \
  -l /export/home/14337185/LW0003.recal.bam \
  -l /export/home/14337185/LW0004.recal.bam \
  -l /export/home/14337185/LW0005.recal.bam \
  -l /export/home/14337185/LW0006.recal.bam \
  -l /export/home/14337185/LW0007.recal.bam \
  -l /export/home/14337185/LW0008.recal.bam \
  -l /export/home/14337185/LW0009.recal.bam \
  -l /export/home/14337185/LW0010.recal.bam \
  -l /export/home/14337185/LW0011.recal.bam \
  -B:dbsnp,vcf /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/dbsnp_132.b37.vcf \
  -o FES.raw.vcf \
  -stand_call_conf 30.0 \
  -stand_emit_conf 10.0 \
  -glm BOTH \
```

The GATK: VariantRecalibrator

```
$ java -Xmx32g -jar /apps/GenomeAnalysisTK-1.1-31-gdc8398e/GenomeAnalysisTK.jar \  
-T VariantRecalibrator \  
-R /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/human_g1k_v37.fasta \  
-B:input,VCF /export/home/14337185/FES.raw.vcf \  
-B:hapmap,VCF,known=false,training=true,truth=true,prior=15.0 /apps/GenomeAnalysisTK-  
1.1-31-gdc8398e/resources/hapmap_3.3.b37.sites.vcf \  
-B:omni,VCF,known=false,training=true,truth=false,prior=12.0 /apps/GenomeAnalysisTK-1.1-  
31-gdc8398e/resources/1000G_omni2.5.b37.sites.vcf \  
-B:dbsnp,VCF,known=true,training=false,truth=false,prior=8.0 /apps/GenomeAnalysisTK-1.1-  
31-gdc8398e/resources/dbsnp_132.b37.vcf \  
-an QD -an HaplotypeScore -an MQRankSum -an ReadPosRankSum -an MQ \  
--maxGaussians 4 \  
--percentBadVariants 0.05 \  
-recalFile /export/home/14337185/FES.vcf.recal \  
-tranchesFile /export/home/14337185/FES.tranches \  
-rscriptFile /export/home/14337185/FES.plots.R \  

```

The GATK: ApplyRecalibration

```
$ java -Xmx32g -jar /apps/GenomeAnalysisTK-1.1-31-gdc8398e/GenomeAnalysisTK.jar \  
-T ApplyRecalibration \  
-R /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/human_g1k_v37.fasta \  
-B:input,VCF /export/home/14337185/FES.raw.vcf \  
--ts_filter_level 99.0 \  
-tranchesFile /export/home/14337185/FES.tranches \  
-recalFile /export/home/14337185/FES.vcf.recal \  
-o /export/home/14337185/FES.vcf \  

```

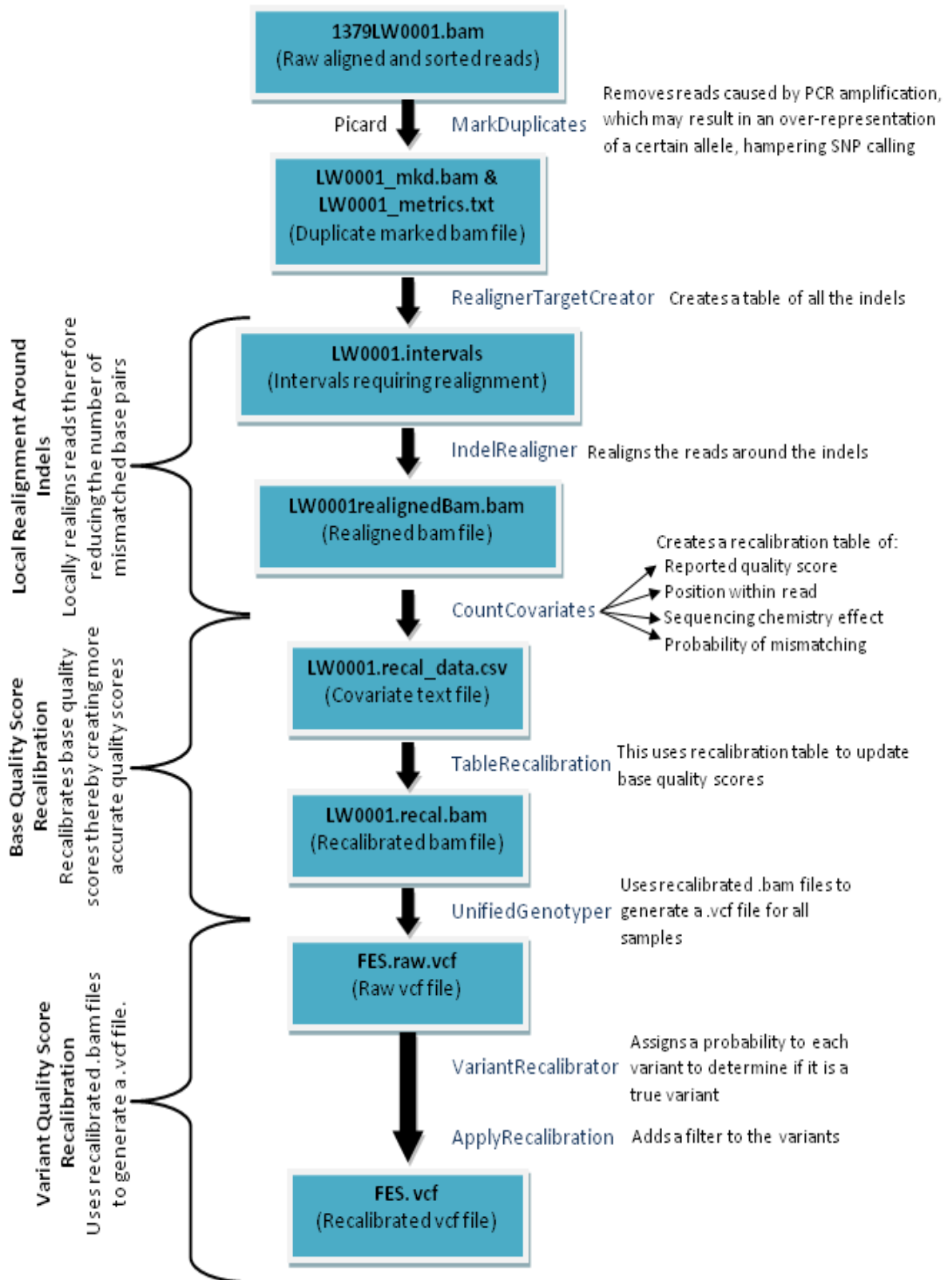


Figure S 4: Variant calling pipeline in the GATK.

APPENDIX 6: Comparison of the non-responder and responder exome data

Commands performed in the Unix shell are indicated in grey font

- The FES.vcf file was used to create variant call format (vcf) files (that record the variant information obtained from NGS data in a standardised format for downstream analyses) containing only the unrelated non-responders and the polymorphic variants in these individuals that passed all the GATK filters, using the GATK's SelectVariants utility. This was repeated to create a vcf file with only the unrelated responders (responders_PASS.vcf):


```
$ java -jar /apps/GenomeAnalysisTK-1.1-31-gdc8398e/GenomeAnalysisTK.jar \
  -R /apps/GenomeAnalysisTK-1.1-31-gdc8398e/resources/human_g1k_v37.fasta \
  -T SelectVariants \
  -B:variant,VCF /export/home/14337185/data/FES.vcf \
  -o /export/home/14337185/SelectVariants/non_reponders_PASS.vcf \
  -sn 1379LW0002 \
  -sn 1379LW0006 \
  -sn 1379LW0009 \
  -sn 1379LW0010 \
  -sn 1379LW0011 \
  -env \
  -ef
```
- The files were then submitted to SeattleSeqAnnotation134 (hg19/GRCh37) for annotation [348].
 - Input files: *VCF SNVs and Indels (both)*
 - All annotation options were selected and the files were *Submitted*
 - The annotated files were decompressed


```
$ gunzip SeattleSeqAnnotation*
```
- Variants occurring in coding regions or splice-sites were selected using pattern scanning with the awk utility:


```
$ awk '$9 ~ /splice-3/' non-responders.txt > non-responders_splice-3.txt
$ awk '$9 ~ /splice-5/' non-responders.txt > non-responders_splice-5.txt
$ awk '$9 ~ /frameshift/' non-responders.txt > non-responders_frameshift.txt
$ awk '$9 ~ /stop-lost/' non-responders.txt > non-responders_stop-lost.txt
$ awk '$9 ~ /stop-gained/' non-responders.txt > non-responders_stop-gained.txt
$ awk '$9 ~ /missense/' non-responders.txt > non-responders_missense.txt
$ awk '$9 ~ /synonymous/' non-responders.txt > non-responders_synonymous.txt
```
- This was repeated for the responders.
- These files were then combined to create single files containing all the coding variants present in the unrelated responders and non-responders:


```
$ cat non-responders_* > non-responders_coding.txt
$ cat responders_* > responders_coding.txt
```

APPENDIX 6

RESPONDER AND NON-RESPONDER COMPARISONS

- Files containing only novel variants were created with awk:


```
$ awk '$11 == "0" { print }' responders_coding.txt > responders_coding_novel.txt
$ awk '$11 == "0" { print }' non_responders_coding.txt > non_responders_coding_novel.txt
```
- Files which contained only the chromosome number and position of the variants, without duplicates, were created for all classes of variation using the SeattleSeqAnnotation files using the cat, cut and sed utilities:


```
$ cat responders_coding.txt | cut -f2,3 | sed '/^\^G/d' | sed 's/\t/:/g' > responders_chr_pos.txt
```
- These files (for the non-responders and responders) were then uploaded onto BioVenn [418] and Venn diagrams were drawn. These files were created as embedded svg files.
- To determine how many of the variants that were not shared between the responders and non-responders were only present in one individual, files were created that only contained the variants that were present in one of the unrelated FES individuals:


```
$ cat FES_PASS.vcf | cut -f1,2,3,4,5,6,7,8,9,10,11,12,13,15,16,17,18,19,20 >
FES_PASS_unrelated.vcf
$ python findSingleton.py > FES_PASS_unrelated_singletons.vcf
```

 Where findSingleton.py was a python script:

```
# Import file data
```

```
id_infilename = "/home/britt/Desktop/vcf_files_singletons/FES_PASS_unrelated.vcf"
```

```
id_infile = open(id_infilename)
```

```
id_lines = id_infile.readlines()
```

```
for id_line in id_lines:
```

```
    id_list = id_line.split()
```

```
    count = 0
```

```
    for item in id_list:
```

```
        if item[0:3] == "0/1":
```

```
            count += 1
```

```
        elif item[0:3] == "1/1":
```

```
            count += 1
```

```
    if count == 1:
```

```
        print id_line
```


APPENDIX 6

RESPONDER AND NON-RESPONDER COMPARISONS

- Thereafter all blank lines were removed using the Unix stream editor, sed:
\$ sed '/^\$/d' FES_PASS_unrelated_singletons.vcf > FES_PASS_unrelated_singletons_noblanks.vcf
- And the header was added:
\$ cat header FES_PASS_unrelated_singletons_noblanks.vcf >
FES_PASS_unrelated_singeltons_noblanks_header.vcf
- Then two separate files were created from this file which contained only the unrelated non-responders in the one file and only the unrelated responders in the other file:
\$ awk '{print \$1,\$2,\$3,\$4,\$5,\$6,\$7,\$8,\$9,\$11,\$14,\$17,\$18,\$19}'
FES_PASS_unrelated_singletons_noblanks_header.vcf > FES_PASS_unrelated_NR.vcf
\$ awk '{print \$1,\$2,\$3,\$4,\$5,\$6,\$7,\$8,\$9,\$10,\$12,\$13,\$15,\$16}'
FES_PASS_unrelatedsingletons_noblanks_header.vcf > FES_PASS_unrelated_R.vcf
- The findSingleton.py python script was run again for each of these files, the blanks were removed and the header was added as previously described.
- These files were then submitted to SeattleSeqAnnotation134 and files with only coding variants were created as described above.
- The number of singleton coding variants in the non-responder and responder groups were then determined as reported in the SeattleSeqAnnotation134 files.

APPENDIX 7: Selection of variants to genotype in the larger cohorts

Commands performed in the Unix shell are indicated in grey font

Variants were selected to be genotyped using the 384-plex GoldenGate Custom Genotyping on VeraCode (Illumina, California). Detailed descriptions for the prioritisation strategy are provided below.

Variants previously associated with antipsychotic response

Previous associations obtained from genome-wide association studies

Using the **HuGE Navigator** (version 2.0), the GWAS integrator was used to search for GWAS associations relating to the search term “antipsychotics” [202]. This search identified 103 variants that were reported to be significantly associated ($P < 1.0 \times 10^{-5}$) with antipsychotic response traits in previous GWAS. The top three-six variants for each reported study or associated trait were prioritised. These 28 variants from the seven reported traits shown below all passed the Illumina ADT analysis and were thus included in the final 384 variant assay. The studies and traits that were identified were as follows:

- Response to iloperidone treatment (QT prolongation) [210]
- Response to antipsychotic therapy (extrapyramidal side effects) [206]
- Treatment response to antipsychotics [208]
- Response to antipsychotic treatment [204]
- Reasoning [207]
- Working memory [207]
- Response to antipsychotic therapy (extrapyramidal side effects) [212]

Previous associations detected in candidate genes

The Pubmed Database [82] was used to search for review articles from the last five years using the search term “antipsychotic pharmacogenetics”. The top 25 genes that were most frequently reported on in these articles were selected for further analyses (refer to Table 2.2 for details on these 25 genes). Of the 25 genes, 8 are already being examined by other projects in our laboratory (*COMT*, *CYP1A2*, *CYP2D6*, *DRD2*, *DRD3*, *DRD4*, *HTR2A*, *SOD2*), thus these genes were excluded from all further analyses.

The articles obtained from the PubMed Database search were subsequently data mined to identify variants that were reported to be associated with antipsychotic response and occurred within the remaining 17 genes. This analysis identified 41 variants, of which 7 were not suitable for genotyping with the BeadXpress Assay. Furthermore, the *HLA* alleles were excluded both due to their complexity and the fact that they have predominantly been associated with clozapine induced agranulocytosis, which was not the drug or trait of interest for this study. Thus 34 variants underwent analyses with Illumina’s ADT. Of these variants, three in *ABCB1* were not supported by the BeadXpress platform due to the fact that they had more than two variants at that site; two in *HTR6* and *LEP* failed due to the failure code 360 (SNP has low score) and two in *SNAP25*, one in *GNB3* and one in *HTR1A* failed

due to the failure code 340 (another marker in the list is closer than 61 nucleotides away). The more frequently referred to *SNAP25* variant was thus included and in the case of *GNB3* and *HTR1A* the functional variants detected in the exome data (see below) were included rather than the previously reported variants. Thus, a total of 22 variants were included in the 384 variant assay. Due to limitations in the size of the assay that we were using, the failed variants were not tagged as the exome variation data was a priority for this study.

Variants identified from the exome data

Functional variants detected in candidate genes

Using SeattleSeqAnnotation134 [348], the vcf files generated by the GATK [326] were annotated as described above. Thereafter the variation present in the 17 prioritised candidate genes described above was identified employing awk:

```
$ awk '§21 ~ /gene_name/' vcf_file_name.txt > gene_name.txt
```

A file containing all the information for the variants present in these candidate genes was created:

```
$ cat gene_name1.txt gene_name2.txt etc > candidate_genes.txt
```

This variation was subsequently prioritised for genotyping by filtering out all variants that did not fall under the following SeattleSeqAnnotation134 functionGVS categories:

- Frameshift
- Stop-lost
- Stop-gained
- Splice-3
- Splice-5
- Missense

For the missense variants to be included, the amino acid changes were required to be predicted to be damaging by either the SIFT [308] or PolyPhen-2 algorithms [356].

These analyses revealed 41 variants. Of these, three variants (*BDNF* Val66Met: rs6265, *LEPR* Lys656Asn: rs8179183, *MTHFR* Ala222Val: rs1901133) had already been included in the assay due to their reported association with antipsychotic response in the literature. After submission to Illumina's ADT, 27 of the 41 variants received failure codes 340 or 360. Of the 27 variants, 25 were in *HLA-DQB1* (22 with the failure code 340 and 3 with the failure code 360). As 22 of the *HLA-DQB1* variants failed due to the failure code 340, it was possible to include six of these in the final assay by removing other variants that were closer than 61 nucleotides. Together these *HLA-DQB1* variants were in perfect LD with seven of the *HLA-DQB1* variants that failed. The remaining two variants that failed were in the *HTT* gene (failure code 360) and one of these variants was in perfect LD with another *HTT* variant that did not fail. Thus, none of the failed SNPs were tagged as the genes were adequately represented by the total of 17 variants that were included in the 384 variant assay.

Variant Annotation, Analysis and Search Tool (VAAST) gene-based analyses

In order to detect the genes that were differentially affected by variants in the responders and non-responders, VAAST [355] was used. To perform these analyses different disease models were used (Figure S5) and only variants that passed the GATK filters were considered.

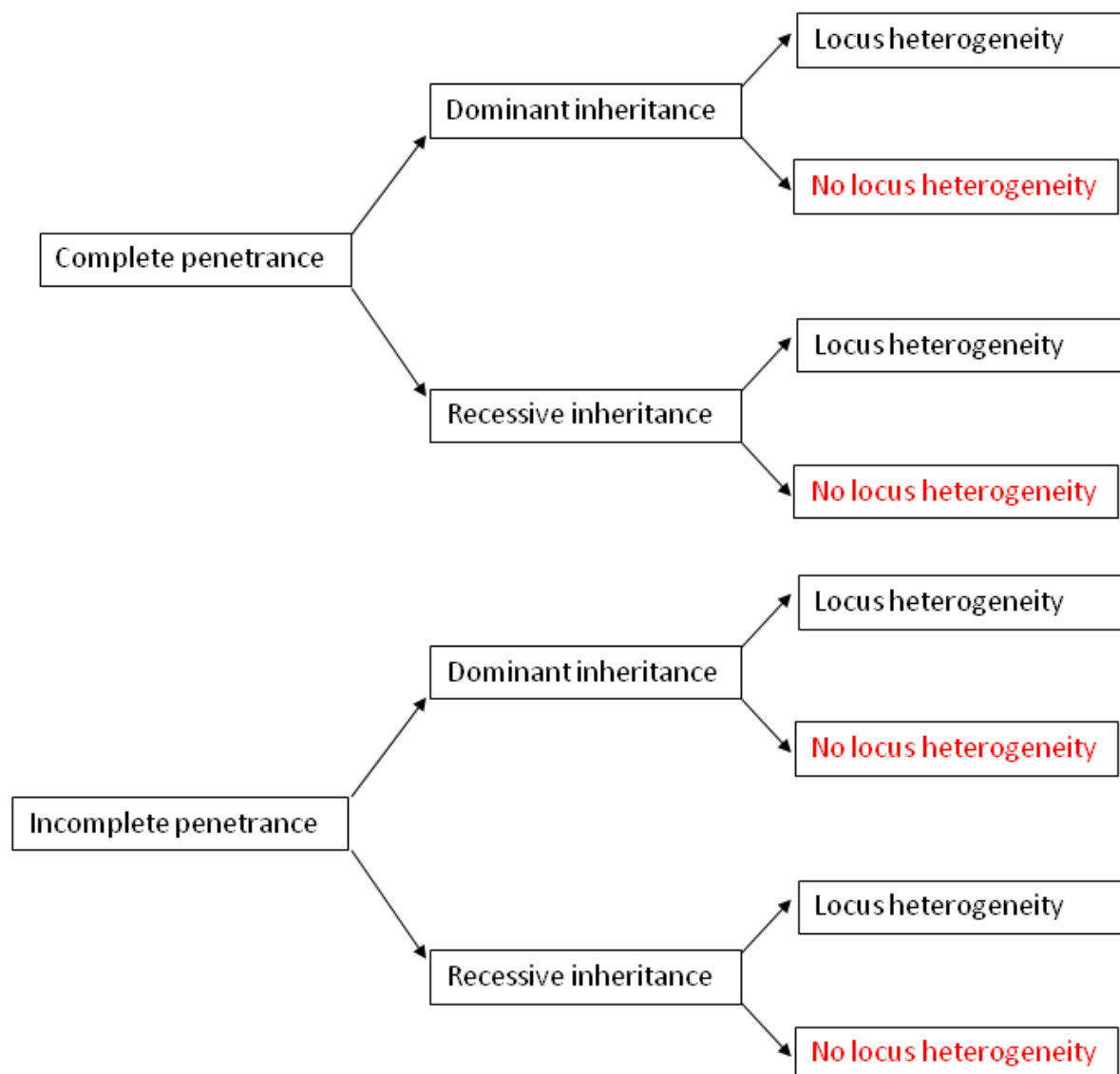


Figure S 5: The models of disease that were used in the VAAST analyses. Those blocks in red revealed no results.

APPENDIX 7

VARIANT PRIORITISATION

To perform the analyses, the `vaast_hsap_chrs_hg19.fa.gz`, `refGene_hg19gff3.gz` files were downloaded.

- Thereafter the genome variation format (gvf) files (variant files that use sequence ontology to describe the data) were created:


```
$ /home/britt/Desktop/VAAST_Code_1.0.3/bin/vaast_tools/vaast_converter -b hg19 -p
/home/britt/Desktop/VAAST/gvf
/home/britt/Desktop/GATK_output/nonfilteredvcf/FES_PASS.vcf
```
- The decoy sequences (`chrGL*`), which have been designed by the Broad institute to remove sequences that do not map well, were removed and `chrMT` was replaced with `chrM`:


```
$ sed '/^chrGL/d' 1379LW0001.gvf > LW0001_nodecoy.gvf
$ sed 's/chrMT/chrM/g' LW0001_nodecoy.gvf > LW0001.gvf
```
- The generic feature format version 3 (gff3) (similar to the gvf files, but with additional features) and gvf files were sorted:


```
$ /apps/VAAST_Code_1.0.3/bin/vaast_tools/vaast_sort_gff
/export/home/14337185/VAAST_data_files/refGene_hg19.nochrM.gff3
/export/home/14337185/gvf/LW0001.gvf /export/home/14337185/gvf/LW0002.gvf
/export/home/14337185/gvf/LW0003.gvf /export/home/14337185/gvf/LW0004.gvf
/export/home/14337185/gvf/LW0005.gvf /export/home/14337185/gvf/LW0006.gvf
/export/home/14337185/gvf/LW0007.gvf /export/home/14337185/gvf/LW0008.gvf
/export/home/14337185/gvf/LW0009.gvf /export/home/14337185/gvf/LW0010.gvf
/export/home/14337185/gvf/LW0011.gvf
```
- The fasta, gff3 and gvf files were indexed:


```
$ /apps/VAAST_Code_1.0.3/bin/vaast_tools/vaast_indexer
/export/home/14337185/VAAST_data_files/vaast_hsap_chrs_hg19.fa
/export/home/14337185/VAAST_data_files/refGene_hg19.nochrM.sorted.gff3
/export/home/14337185/gvf/LW0001.sorted.gvf
/export/home/14337185/gvf/LW0002.sorted.gvf
/export/home/14337185/gvf/LW0003.sorted.gvf
/export/home/14337185/gvf/LW0004.sorted.gvf
/export/home/14337185/gvf/LW0005.sorted.gvf
/export/home/14337185/gvf/LW0006.sorted.gvf
/export/home/14337185/gvf/LW0007.sorted.gvf
/export/home/14337185/gvf/LW0008.sorted.gvf
/export/home/14337185/gvf/LW0009.sorted.gvf
/export/home/14337185/gvf/LW0010.sorted.gvf
/export/home/14337185/gvf/LW0011.sorted.gvf
```
- The gvf files were annotated using the variant annotation tool (VAT). Additionally, the default chunk size were changed to 263000000 (the size of the biggest chromosome):


```
$ /apps/VAAST_Code_1.0.3/bin/VAT -f
/export/home/14337185/VAAST_data_files/refGene_hg19.gff3 -a
/export/home/14337185/VAAST_data_files/vaast_hsap_chrs_hg19.fa -c 263000000 -g female -
b hg19 /export/home/14337185/gvf/LW0001.sorted.gvf >
/export/home/14337185/gvf/LW0001.vat.gvf
```

- This was repeated for all samples by changing the sample names and the gender, where: LW0001, LW0002, LW0003, LW0005, LW0011 are female and LW0004, LW0006, LW0007, LW0008, LW0009, LW0010 are male. These *.vat.gvf files were then all indexed as described previously.
- The following condenser (cdr) files (which contain the merged information from the gvf files) were created:

responders.cdr

```
$ /apps/VAAST_Code_1.0.3/bin/VST -o 'U(0, 2, 3, 6, 7)' -b hg19
/export/home/14337185/gvf/LW0001.vat.gvf /export/home/14337185/gvf/LW0002.vat.gvf
/export/home/14337185/gvf/LW0003.vat.gvf /export/home/14337185/gvf/LW0004.vat.gvf
/export/home/14337185/gvf/LW0005.vat.gvf /export/home/14337185/gvf/LW0006.vat.gvf
/export/home/14337185/gvf/LW0007.vat.gvf /export/home/14337185/gvf/LW0008.vat.gvf
/export/home/14337185/gvf/LW0009.vat.gvf /export/home/14337185/gvf/LW0010.vat.gvf
/export/home/14337185/gvf/LW0011.vat.gvf > /export/home/14337185/cdr/responders.cdr
```

non responders with l.cdr

```
$ /apps/VAAST_Code_1.0.3/bin/VST -o 'U(l(4, 5),U(1, 8, 9, 10))' -b hg19
/export/home/14337185/gvf/LW0001.vat.gvf /export/home/14337185/gvf/LW0002.vat.gvf
/export/home/14337185/gvf/LW0003.vat.gvf /export/home/14337185/gvf/LW0004.vat.gvf
/export/home/14337185/gvf/LW0005.vat.gvf /export/home/14337185/gvf/LW0006.vat.gvf
/export/home/14337185/gvf/LW0007.vat.gvf /export/home/14337185/gvf/LW0008.vat.gvf
/export/home/14337185/gvf/LW0009.vat.gvf /export/home/14337185/gvf/LW0010.vat.gvf
/export/home/14337185/gvf/LW0011.vat.gvf >
/export/home/14337185/cdr/non_responders_with_l.cdr
```

- The following analyses were performed using the variant annotation, analysis and search tool (VAAST) as shown in Figure S5:

Complete penetrance, dominant inheritance and locus heterogeneity:

```
$ /home/britt/Desktop/VAAST_Code_1.0.3/bin/VAAST -gp 4260 -m lrt -ref -k -pnt c -iht d -o
/home/britt/Desktop/VAAST/analysis_c_d_y
/home/britt/Desktop/VAAST/VAAST_data_files/refGene_hg19.gff3
/home/britt/Desktop/VAAST/cdr/responders.cdr
/home/britt/Desktop/VAAST/cdr/non_responders_with_l.cdr
```

Complete penetrance, dominant inheritance and no locus heterogeneity:

```
$ /home/britt/Desktop/VAAST_Code_1.0.3/bin/VAAST -gp 4260 -m lrt -ref -k -pnt c -iht d -lh n
-o /home/britt/Desktop/VAAST/analysis_c_d_n
/home/britt/Desktop/VAAST/VAAST_data_files/refGene_hg19.gff3
/home/britt/Desktop/VAAST/cdr/responders.cdr
/home/britt/Desktop/VAAST/cdr/non_responders_with_l.cdr
```

Complete penetrance, recessive inheritance and locus heterogeneity:

```
$ /home/britt/Desktop/VAAST_Code_1.0.3/bin/VAAST -gp 4260 -m lrt -ref -k -pnt c -iht r -o
/home/britt/Desktop/res_nonres_l/analysis_c_r_y
/home/britt/Desktop/VAAST/VAAST_data_files/refGene_hg19.gff3
/home/britt/Desktop/VAAST/cdr/responders.cdr
/home/britt/Desktop/VAAST/cdr/non_responders_with_l.cdr
```

Complete penetrance, recessive inheritance and no locus heterogeneity:

```
$ /home/britt/Desktop/VAAST_Code_1.0.3/bin/VAAST -gp 4260 -m lrt -ref -k -pnt c -iht r -lh n
-o /home/britt/Desktop/res_nonres_l/analysis_c_r_n
/home/britt/Desktop/VAAST/VAAST_data_files/refGene_hg19.gff3
/home/britt/Desktop/VAAST/cdr/responders.cdr
/home/britt/Desktop/VAAST/cdr/non_responders_with_l.cdr
```

Incomplete penetrance, dominant inheritance and locus heterogeneity:

```
$ /home/britt/Desktop/VAAST_Code_1.0.3/bin/VAAST -gp 4260 -m lrt -ref -k -iht d -o
/home/britt/Desktop/res_nonres_l/analysis_i_d_y
/home/britt/Desktop/VAAST/VAAST_data_files/refGene_hg19.gff3
/home/britt/Desktop/VAAST/cdr/responders.cdr
/home/britt/Desktop/VAAST/cdr/non_responders_with_l.cdr
```

Incomplete penetrance, dominant inheritance and no locus heterogeneity:

```
$ /home/britt/Desktop/VAAST_Code_1.0.3/bin/VAAST -gp 4260 -m lrt -ref -k -iht d -lh n -o
/home/britt/Desktop/res_nonres_l/analysis_i_d_n
/home/britt/Desktop/VAAST/VAAST_data_files/refGene_hg19.gff3
/home/britt/Desktop/VAAST/cdr/responders.cdr
/home/britt/Desktop/VAAST/cdr/non_responders_with_l.cdr
```

Incomplete penetrance, recessive inheritance and locus heterogeneity:

```
$ /home/britt/Desktop/VAAST_Code_1.0.3/bin/VAAST -gp 4260 -m lrt -ref -k -iht r -o
/home/britt/Desktop/res_nonres_l/analysis_i_r_y
/home/britt/Desktop/VAAST/VAAST_data_files/refGene_hg19.gff3
/home/britt/Desktop/VAAST/cdr/responders.cdr
/home/britt/Desktop/VAAST/cdr/non_responders_with_l.cdr
```

Incomplete penetrance, recessive inheritance and no locus heterogeneity:

```
$ /home/britt/Desktop/VAAST_Code_1.0.3/bin/VAAST -gp 4260 -m lrt -ref -k -iht r -lh n -o
/home/britt/Desktop/res_nonres_l/analysis_i_r_n
/home/britt/Desktop/VAAST/VAAST_data_files/refGene_hg19.gff3
/home/britt/Desktop/VAAST/cdr/responders.cdr
/home/britt/Desktop/VAAST/cdr/non_responders_with_l.cdr
```

APPENDIX 7

VARIANT PRIORITISATION

No results were obtained when “no locus heterogeneity” was included. Thus, only the remaining four disease models could be used for variant selection. From each of these analyses the top 20 genes were identified and the variation present in these genes was prioritised utilising the following criteria:

- All variants with a VAAST likelihood ratio greater than 0
- The SeattleSeqAnnotation134 annotation was used, as described above, to identify additional functional variation present in these genes. These variants were identified by filtering out all variants that do not fall under the following SeattleSeqAnnotation134 functionGVS categories:
 - Frameshift
 - Stop-lost
 - Stop-gained
 - Splice-3
 - Splice-5
 - Missense
 - For the missense variants to be included, the amino acid changes were required to be predicted to be damaging by both the SIFT [308] or PolyPhen-2 [356] algorithms.
- As an additional filter, all variants that were monomorphic, had low coverage and occurred on the X chromosome as heterozygotes in males were excluded.

These analyses identified 162 variants in 55 genes. After Illumina’s ADT analysis, it was found that twelve variants failed due to the failure code 360. Of these variants, six (in the genes *GLIS1*, *ZC3H3*, *ANKRD53*, *CHFR* and *ZNF470*) were tagged by other variants (refer to the paragraph below for details on how tagging SNPs were identified). For the remaining six variants (in the genes *N4BP3*, *NOV*, *SIPA1L2*, *ADO* and *KRT5*), no tagging variants were identified. An additional 22 variants failed due to the failure code 340. To determine which variant from each of these 11 pairs of variants would be included, the following strategy was used: (i) determine which of the variants in the pair is better tagged by another variant and then include the other variant in the assay; (ii) choose the variant which is validated by VAAST in the assay; (iii) choose the variant which is predicted by SIFT and PolyPhen to be more damaging. From the 11 variants that remained, only 1 (in *GPRIN2*) was not tagged by any other variants. This left 139 variants that could be included in the assay. The remaining variants were tagged by 19 SNPs. Thus a total of 158 variants were included in the assay to represent the VAAST gene analysis. The variants that were not able to be genotyped or tagged by other variants were genotyped with the use of PCR-RFLP/TaqMan analyses in the South African FES cohort.

To account for the variants that were not able to be genotyped, haplotype data from the exome sequence individuals and the 1000 Genomes Project were used. Haplotype data for the three African populations (ASW, LWK, YRI) available on the 1000 Genomes Project browser [349] were loaded onto Haploview v4.0 [419] and the default settings were used to identify variants that were in LD with the variants that failed the ADT ($r^2 > 0.8$, $LOD > 3$). Tagging variants were chosen as follows: (i) if a variant was shown to be in perfect LD in the eleven exome sequence individuals, but this pattern of LD is not confirmed by the 1000 Genomes Project data, the confidence of this tagging is not high. Therefore, all variants showing this pattern should be genotyped; (ii) if a variant is in LD with the variant of interest in all three 1000 Genomes Project populations, this variant is considered a good tagging SNP and thus only this variant is included; (iii) if a variant is in perfect LD in the eleven exome sequence individuals and this pattern is observed in at least one 1000 Genomes Project population, this variant

is considered a good tagging SNP and thus only this variant is included; and (iv) if a variant is not present in the exome data and LD is observed in less than three 1000 Genomes Project populations, where possible, at least two variants are required to tag the SNP.

Variant Annotation, Analysis and Search Tool (VAAST) variant-based analyses

To determine which variants occurred in three or more non-responders and none of the responders and *vice versa*, the following analyses were performed in VAAST [355], considering only variants that passed the GATK filters:

```
$ /apps/VAAST_Code_1.0.3/bin/VST -o 'C(S(">3",1,4,5,8,9,10),S(">0",0,2,3,6,7))' -b hg19
/export/home/14337185/gvf/LW0001.vat.gvf /export/home/14337185/gvf/LW0002.vat.gvf
/export/home/14337185/gvf/LW0003.vat.gvf /export/home/14337185/gvf/LW0004.vat.gvf
/export/home/14337185/gvf/LW0005.vat.gvf /export/home/14337185/gvf/LW0006.vat.gvf
/export/home/14337185/gvf/LW0007.vat.gvf /export/home/14337185/gvf/LW0008.vat.gvf
/export/home/14337185/gvf/LW0009.vat.gvf /export/home/14337185/gvf/LW0010.vat.gvf
/export/home/14337185/gvf/LW0011.vat.gvf >
/export/home/14337185/cdr/nonresponders_shared.cdr
```

```
$ /apps/VAAST_Code_1.0.3/bin/VST -o 'C(S(">3",0,2,3,6,7),S(">0",1,4,5,8,9,10))' -b hg19
/export/home/14337185/gvf/LW0001.vat.gvf /export/home/14337185/gvf/LW0002.vat.gvf
/export/home/14337185/gvf/LW0003.vat.gvf /export/home/14337185/gvf/LW0004.vat.gvf
/export/home/14337185/gvf/LW0005.vat.gvf /export/home/14337185/gvf/LW0006.vat.gvf
/export/home/14337185/gvf/LW0007.vat.gvf /export/home/14337185/gvf/LW0008.vat.gvf
/export/home/14337185/gvf/LW0009.vat.gvf /export/home/14337185/gvf/LW0010.vat.gvf
/export/home/14337185/gvf/LW0011.vat.gvf > /export/home/14337185/cdr/responders_shared.cdr
```

- These two files were then combined using the following command:
\$ cat nonresponders_shared.cdr responders_shared.cdr > shared.cdr
- One was subtracted from all chromosome positions for nucleotide deletions (to ensure that the cdr and vcf positions for the indels correlate, as vcf deletions are padded).
- Thereafter the analyses described below were performed to create a vcf file containing the variants that were present in the cdr file to allow for submission SeattleSeqAnnotation 134.

For the cdr file:

- A file with only columns 1 and 2 (chromosome and position), no tabs and no 'chr' was created:
\$ cat shared.cdr | cut -f1,2 | sed 's/\t/,/g' | sed 's/chr//' > FEScdrcommanochr.txt
- A file with the remaining columns (3-8) was created:
\$ cat shared.cdr | cut -f3,4,5,6,7,8 shared.cdr > FEScdr3_8.cdr
- These two files were combined:
\$ paste -d' ' FEScdrcommanochr.txt FEScdr3_8.cdr > FEScdrall.txt

For the vcf file:

- A file with only columns 1 and 2 (chromosome and position) and no tabs was created:
\$ cat FES_PASS.vcf | cut -f1,2 | sed 's/\t/,/g' > FESvcfcomma.txt
- A file with the remaining columns (3-20) was created:
\$ cat FES_pass.vcf | cut -f3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20 FES_pass.vcf > FESvcf3_20.txt

APPENDIX 7

VARIANT PRIORITISATION

- These two files were combined:
\$ paste -d' ' FESvcfcomma.txt FESvcf3-20.txt > FESvcfall.txt
- All the lines that were shared between the cdr and vcf files were printed:
\$ awk 'FNR==NR{ a[\$1]=\$0;next } (\$1 in a)' FEScdrall.txt FESvcfall.txt > combined.txt
- The number of lines in each of the files were counted:
- \$ wc -l shared.cdr
\$ wc -l combined.txt
- The number of header lines was subtracted:
\$ grep -c "^#" shared.cdr
\$ grep -c "^#" combined.txt
- There were two extra lines in the FESvcfall.txt file that correlated to two chromosome positions that had both an indel and a SNV at this position. After determining which of these variants the variant of interest was, the extra variant was deleted.
- To ensure that the combined .txt file was in the same format as a vcf file, the data was pasted into excel and the data was moved to columns using the “text to columns” function in Microsoft Excel (2007).
- Thereafter the vcf header was added to the combined.txt file
\$ cat header.txt combined.txt > Seattleseqfreq.vcf

The Seattleseqfreq.vcf file was then submitted to SeattleSeqAnnotation134 [348] and the variants were filtered based on the following criteria:

- Variants that had genotype information for at least 10 of the individuals.
- Variants occurring within miRNAs
- Variants falling under the following SeattleSeqAnnotation134 functionGVS or functionGVS categories:
 - Frameshift
 - Stop-lost
 - Stop-gained
 - Splice-3
 - Splice-5
 - Missense

For the missense variants to be included, the amino acid changes were required to be predicted to be damaging by either the SIFT or PolyPhen algorithms

These analyses detected 56 variants and 49 of these variants passed the Illumina ADT criteria. Four of the variants that failed the ADT analyses (in *UPP2* and *SLC39A4*) failed due to the failure code 340. Due to the fact that all of these variants were considered important, in each case one variant was included in the BeadXpress Assay and the other variant was manually genotyped with the use of PCR-RFLP (*SLC39A4*) or a custom TaqMan assay (*UPP2*). A further three variants failed (in *PLEKHG3*, *TCF25* and *NDOR1*) due to the failure code 360, however five tag SNPs for all of these variants were identified, as previously described. Thus a total of 56 variants were included in the 384 variant assay.

APPENDIX 8: Genotyping protocols for prioritised variants

Polymerase chain reaction amplification

Variants that were not included in the Illumina BeadXpress Assay were genotyped by means of PCR-RFLP genotyping. Primers for the amplification of the regions containing the variants of interest were designed with the use of the PrimerQuest [420], OligoAnalyzer 3.0 [421] and PrimerBLAST [422] computational tools. All reference sequences were obtained from Ensembl [395] (refer to Table S7 for primer sequences). All PCR amplification reactions were prepared to a final volume of 25 μ l, containing 20 ng of gDNA (or a 1/100 dilution of PCR product for nested PCR reactions), a final concentration of 1X buffer, 1.5 mM MgCl₂, 0.4 mM dNTPs, 0.4 μ M of both the forward and reverse primer and 0.5 U BIOTAQ™ Polymerase (Bioline, UK). In addition, the PCR for the amplification of the *ADO* region utilised a final concentration of 1 M Betaine (Sigma-Aldrich (Pty) Ltd, Aston Manor, South Africa); while the PCR reactions for *N4BP3* and *NOV* utilised 5% Dimethyl sulfoxide (DMSO) (Sigma-Aldrich (Pty) Ltd, Aston Manor, South Africa). All amplification cycles reactions consisted of an initial denaturation of 94°C for 3 min; followed by 40 cycles (or 25 cycles for nested PCR reactions) of denaturation at 94°C for 15 sec, annealing for 15 sec at the temperatures given in Table S7, and extension at 72°C for 30 seconds; after which the cycles were concluded with a final extension step at 72°C for 5 min. The resulting amplicons were subsequently examined by loading 5 μ l of PCR product and cresol loading buffer each into 1% (w/v) ethidium bromide-stained agarose gels, which then underwent gel electrophoresis in 1X TBE gel electrophoresis buffer, at 120 V. The products were subsequently analysed under UV light at 260 nm and HyperLadder IV (Bioline, UK) was used as a molecular weight marker.

Sanger sequence verification

The presence of the variants that were to be genotyped by means of PCR-RFLP was verified with the use of Sanger sequencing of selected individuals testing positive for the variants of interest. After amplification, SureClean (Bioline, UK) was used to purify the resulting PCR products, according to the specified manufacturer's protocol. Thereafter, the NanoDrop® ND-100 (Nanodrop Technologies Inc., Delaware, USA) was used to determine the concentration of the purified product by measuring the absorbance at 260 nm and the appropriate dilutions for the sequencing reaction were prepared. Sequencing reactions utilised Big Dye v3.1 sequencing chemistry with the addition of Half Dye Mix (Bioline, UK) according to the manufacturer's recommendations. The sequencing cycle reactions were performed as follows: initial denaturation of 94°C for 5 min, followed by 25 cycles of denaturation at 94°C for 10 min, annealing at 55°C for 10 min and extension at 60°C for 4 min. After the sequencing reaction was complete, a purification cycle of 98°C for 5 min and 25°C for 10 min was performed by means of the addition of 0.2% (w/v) SDS. Subsequent capillary electrophoresis was performed by the Central Analytical Facility of Stellenbosch University on a 3130XI Genetic Analyzer, according to the manufacturer's protocol (Applied Biosystems™, California, USA). The resulting sequence data was aligned to the reference sequence of interest using BioEdit v7.0.9.0 [423] in order to verify the presence of the variant of interest.

Table S 7: Primer sequences for amplification of regions containing the variants that were not genotyped by the Illumina BeadXpress Assay

Gene	Region	Primer Name	Primer Sequence (5'-3')	Tm (°C)	Product Size (bp)
<i>ADO</i>	Exon 1	<i>ADOF</i>	CGC GTG GCT GCT GAG GTT GGC GG	68	305
		<i>ADOmR</i>	AAG CCC GGC TGC ATC GGC GCC cCC		
<i>GRPIN2</i>	Exon 1	<i>GRPIN2F</i> ¹	CCT GGC TCC TGA GGA TGA GAC TTC	60	443
		<i>GPRIN2R</i> ¹	CCA AGT CAT TGG CTG AGG TCA TGG TCC		
<i>GRPIN2</i>	Exon 1	<i>GRPIN2F</i>	CCT GGC TCC TGA GGA TGA GAC TTC	55	229
		<i>GPRIN2mR</i>	AGG TAG GGC ATG GCA GCA GCC tgC A		
<i>KRT5</i>	Exon 2	<i>KRT5mF</i>	GCA GGA GCA GGG CAC CAc GAC T	68	283
		<i>KRT5R</i>	GTC CAT GGA AGG TAT ATC CTC CCA GCC CC		
<i>N4BP3</i>	Exon 2	<i>N4BP3F</i>	TGT GAG AGC ATC AGG AGG TAG AGC A	58	638
		<i>N4BP3R</i>	TGA GGC ACA GAA CAG GAC TCC A		
<i>NOV</i>	Exon 2	<i>NOVF</i> ²	CGA GCA GTG CCA ATC TAC AGC G	68	548
		<i>NOVR</i> ²	CTT AGC TGC AGG AGA AGA GGT CAA AC		
<i>NOV</i>	Exon 2	<i>NOVmF</i> ³	AGC GCT GCC CTC CCC AGT GCC CGG aCC	68	259
		<i>NOVR</i> ³	CTT AGC TGC AGG AGA AGA GGT CAA AC		
<i>SIPA1L2</i>	Exon 14	<i>SIPA1L2F</i>	CAG CAG CAA CAC GCT CTC CAG CAA CAC C	58	320
		<i>SIPA1L2mR</i>	TTC CGC AGC ACT GCC GGC caA GAT G		
<i>SLC39A4</i>	Exon 6	<i>SLC39A4F</i>	ACT ACA TCC TGC AGA CCT TCC TGA G	65	350
		<i>SLC39A4R</i>	AGG TTC TCA AAC AGG AAG AAG GCG		

F: Forward primer; R: Reverse primer; m: Mutagenic primer; Lowercase bold letters: Mutagenic bases; 1: External primers used for Sanger sequencing; 2: External primers for nested PCR; 3: Internal primers for nested PCR

Restriction fragment length polymorphism analyses

All variants whose presence was verified by Sanger sequencing confirmation underwent RFLP genotyping analyses, with the exception of rs3812082 in *N4BP3* which was genotyped using Sanger sequencing as described above. The relevant restriction enzymes were identified with the use of RestrictionMapper [424] and in cases where no suitable restriction enzyme were available, mutagenic primers were designed to introduce appropriate restriction enzyme recognition sites. To ensure that the restriction enzyme digestions were reliable, positive controls from the exome sequence data were utilised. After amplification reactions were complete, all restriction enzyme digest reactions were prepared to a final volume of 20 µl and the reaction mix was prepared according to the manufacturers protocol utilising 10 µl of PCR product (refer to Table S8 for RFLP specifications). The fragments from the resulting digests were then examined by loading 10 µl of restriction enzyme digest and cresol loading buffer each into 3% (w/v) ethidium bromide-stained agarose gels, which then underwent gel electrophoresis in 1X TBE gel electrophoresis buffer, at 80 V. The products were subsequently analysed under UV light at A260 nm and HyperLadder V (Bioline, UK) was used as a molecular weight marker.

TaqMan® SNP genotyping

Genotyping of rs11368509 in *UPP2* was performed by means of a custom TaqMan SNP genotyping assay (Applied Biosystems™, California, USA). The genotyping was performed using a reaction mix containing 1x TaqMan Genotyping Master Mix, 1x custom TaqMan SNP assay and 15 ng of gDNA, prepared to a final volume of 10 µl. The amplification was performed in a StepOnePlus™ Real-Time PCR System (Applied Biosystems™, California, USA) utilising a reaction cycle which consisted of an initial denaturation of 95°C for 10 min; followed by 40 cycles of denaturation at 95°C for 15 sec and annealing and extension at 60°C for 1 min. After amplification, allelic discrimination was performed using the Sequence Detection System Software (Applied Biosystems™, California, USA) as described in the manufacturer's protocol.

Illumina BeadXpress Assay genotyping

All samples were obtained from whole blood and gDNA was previously extracted using the Miller *et al.* [354] protocol. The concentrations of the gDNA samples were determined with the use of a NanoDrop® ND-100 (Nanodrop Technologies Inc., Delaware, USA) at an absorbance of 260 nm. The samples were subsequently diluted to 50 ng/µl in TE. A final volume of 10 µl was pipetted into four 96-well plates, which were sealed with optical adhesive film and frozen in preparation for shipping. The samples were subsequently shipped (as previously described for the gDNA samples utilised for the exome sequencing) to the University of Utah genomics core research facility [425] for genotyping with the VeraCode® technology. The Illumina BeadXpress GoldenGate® assay was ordered from Illumina and sent directly to the University of Utah Genomics Core Research Facility.

Table S 8: RFLP specifications

Enzymes, buffers and additives were supplied by New England Biolabs Inc., Beverly, USA

Gene	Variant	Chromosome	Position	Restriction Enzyme	Temperature (°C) and Additives	Incubation Time (hours)	Genotype	Size of Fragments (bp)	Primer set
<i>ADO</i>	rs2236295	10	64564892	<i>PvuII</i>	37	2	GG	305	<i>ADOF</i> <i>ADOmR</i>
							GT	305, 236, 69	
							TT	236, 69	
<i>ADO</i>	rs10995311	10	64564934	<i>SmaI</i>	25	2	CC	206, 75, 24	<i>ADOF</i> <i>ADOmR</i>
							CG	230, 206, 75, 24	
							GG	230, 75	
<i>N4BP3</i>	rs3812082*	5	177547336	NA	NA	NA	NA	<i>N4BP3F</i> <i>N4BP3R</i>	
<i>NOV</i>	rs2279112	8	120429024	<i>AgeI</i>	25, BSA	16	AA	259	<i>NOVmF</i> <i>NOVR</i>
							AG	259, 234, 25	
							GG	234, 25	
<i>SIPA1L2</i>	rs2275307	1	232574921	<i>BstXI</i>	37	2	AA	298, 22	<i>SIPA1L2F</i> <i>SIPA1L2mR</i>
							AG	320, 298, 22	
							GG	320	
<i>SLC39A4</i>	rs75920625	8	145639654	<i>HaeIII</i>	37, BSA	2	AA	230, 70, 50	<i>SLC39A4F</i> <i>SLC39A4R</i>
							AG	230, 158, 72, 70, 50	
							GG	158, 72, 70, 50	

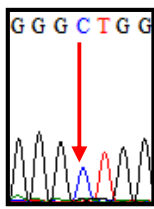
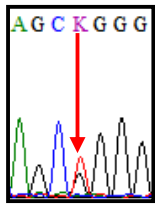
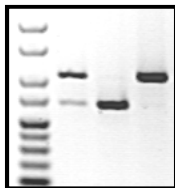
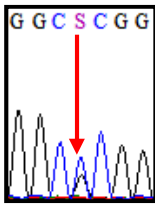
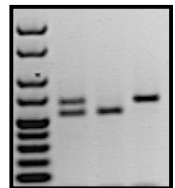
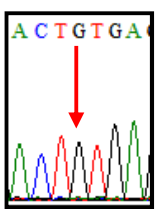
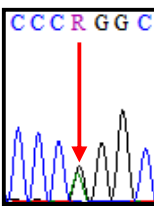
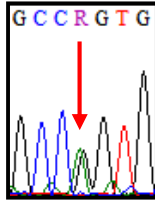
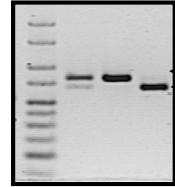
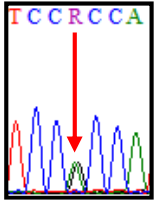
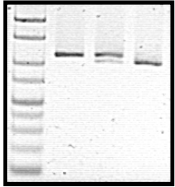
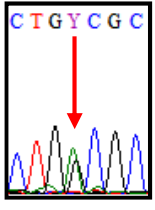
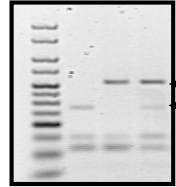
*: This variant was genotyped using Sanger sequencing, F: Forward primer, R: Reverse primer, m: Mutagenic primer

APPENDIX 9: Genotyping results for prioritised variants

PCR-RFLP and TaqMan Genotyping

Sanger sequence verification revealed that two variants (a novel frameshift variant in *GPRIN2* and rs147498164 in *KRT5*) were not present and were thus detected in the exome data as a result of sequencing artefacts. The presence of the remaining variants was confirmed and RFLP or TaqMan SNP genotyping was successfully performed for the genotyping of the remaining samples (Table S9 and Figure S6).

Table S 9: Sanger sequencing conformation and PCR-RFLP genotyping of variants that were not genotyped with the Illumina BeadXpress Assay

<i>GRPIN2</i>	<i>ADO</i>	<i>ADO</i>	<i>KRT5</i>
Novel, Frameshift	rs2236295, G25T	rs10995311, P39A	rs147498164, V211M
 <p>Not present</p>	 <p>HLV TT GT GG</p>  <p>← 305bp ← 236bp</p> <p><i>PvuII</i></p>	 <p>HLV CG CC GG</p>  <p>← 230bp ← 206bp</p> <p><i>SmaI</i></p>	 <p>Not present</p>
<i>N4BP3</i>	<i>NOV</i>	<i>SIPA1L2</i>	<i>SLC39A4</i>
rs3812082, R163G	rs2279112, R42G	rs2275307, T1322A	rs75920625, T356A
 <p>Genotyped with Sanger sequencing</p>	 <p>HLV GA AA GG</p>  <p>← 259bp ← 234bp</p> <p><i>AgeI</i></p>	 <p>HLV AA AG GG</p>  <p>← 319bp ← 298bp</p> <p><i>BstXI</i></p>	 <p>HLV GG AA AG</p>  <p>← 230bp ← 158bp</p> <p><i>HaeIII</i></p>

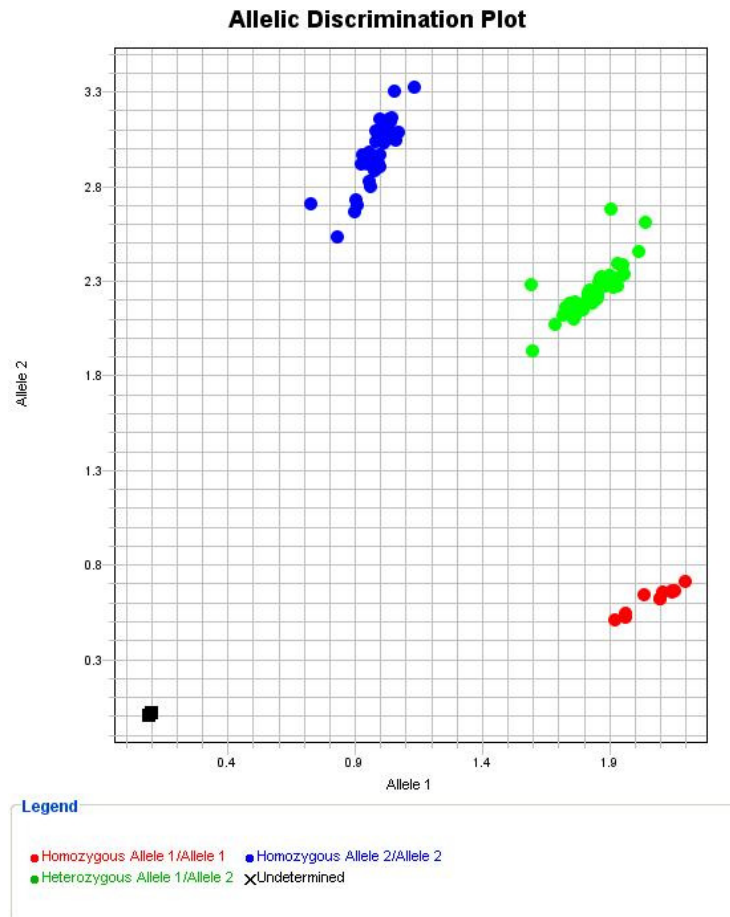


Figure S 6: TaqMan SNP genotyping results for rs11368509 in *UPP2*

Illumina BeadXpress genotyping results

The generated genotyping results were critically examined in GenomeStudio (Illumina, California, USA) to determine which variants should be excluded from the statistical analyses. Table S10 lists all the SNPs that were excluded from the analyses and the corresponding exclusion criteria, while Figure S7 provides examples of the clusters viewed in GenomeStudio.

Table S 10: SNPs that failed the Illumina BeadXpress genotyping

SNP	Exclusion criteria	Type of SNP	Gene
seq-rs2919308	Less than 90% success	AIM	NA
seq-rs10194455	Less than 90% success	AIM	NA
seq-rs10856819	Less than 90% success	AIM	NA
seq-rs12132696	Less than 90% success	AIM	NA
seq-rs7584977	Overlapping clusters	AIM	NA
seq-rs114298106	Less than 90% success	Frequency	<i>C7orf71</i>
seq-rs17561	Less than 90% success	Frequency	<i>IL1A</i>
seq-rs61732484	Less than 90% success	Frequency	<i>OR2T10</i>
seq-rs6710480	Less than 90% success	Frequency	<i>UPP2</i>
seq-rs74730740	Less than 90% success	Frequency	<i>OR4D2</i>
seq-rs28664620	Less than 90% success	Frequency	<i>OR2T10</i>
DNMBP	Monomorphic	Frequency	<i>DNMBP</i>
seq-rs3795789	Monomorphic	Frequency	<i>OBSCN</i>
seq-rs7968606	Less than 90% success	GWAS	NA
seq-rs888219	Overlapping clusters	GWAS	NA
seq-rs8009244	Overlapping clusters	LD VAAST with rs2274271	<i>DLGAP5</i>
seq-rs77099184	Less than 90% success	LD VAAST with rs61729867	<i>FREM3</i>
seq-rs3127821	Less than 90% success	LD VAAST with rs67531787	<i>GPRIN2</i>
seq-rs3127823	Less than 90% success	LD VAAST with rs67531787	<i>GPRIN2</i>
HTR2C	Less than 90% success	PubMed Exome	<i>HTR2C</i>
seq-rs1049056	Less than 90% success	PubMed Exome	<i>HLA-DQB1</i>
seq-rs1049066	Less than 90% success	PubMed Exome	<i>HLA-DQB1</i>
seq-rs1130432	Less than 90% success	PubMed Exome	<i>HLA-DQB1</i>
seq-rs1140318	Less than 90% success	PubMed Exome	<i>HLA-DQB1</i>
seq-rs41544112	Less than 90% success	PubMed Exome	<i>HLA-DQB1</i>
seq-rs61733142	Less than 90% success	PubMed Exome	<i>HTR6</i>
seq-rs9274395	Less than 90% success	PubMed Exome	<i>HLA-DQB1</i>
HTT	Monomorphic	PubMed Exome	<i>HTT</i>
seq-rs5442	Overlapping clusters	PubMed Exome	<i>GNB3</i>
seq-rs701564	Overlapping clusters	PubMed Exome	<i>HLA-DQB1</i>
rs1800629	Less than 90% success	PubMed Known	<i>TNF-alpha</i>
DLGAP5	Less than 90% success	VAAST	<i>DLGAP5</i>
seq-rs1281013	Less than 90% success	VAAST	<i>C1orf127</i>
seq-rs9892256	Less than 90% success	VAAST	<i>DNAH9</i>
seq-rs78610683	Less than 90% success	VAAST	<i>PLEKHG3</i>
EHMT2_2	Less than 90% success	VAAST	<i>EHMT2</i>
N4BP3	Less than 90% success	VAAST	<i>N4BP3</i>
seq-rs3796100	Less than 90% success	VAAST	<i>ANKRD53</i>
seq-rs71520524	Less than 90% success	VAAST	<i>ZC3H3</i>
seq-rs72780221	Less than 90% success	VAAST	<i>GPRIN2</i>
DNHD1_2	Less than 90% success	VAAST	<i>DNHD1</i>
GPRIN2	Monomorphic	VAAST	<i>GPRIN2</i>

APPENDIX 9

GENOTYPING RESULTS

seq-rs34110867	Monomorphic	VAAST	<i>SIPA1L2</i>
RBMXL3_1	Overlapping clusters	VAAST	<i>LRCH2,RBMXL3</i>
seq-rs11204658	Overlapping clusters	VAAST	<i>GPRIN2</i>
seq-rs1770984	Overlapping clusters	VAAST	<i>ASPG</i>
seq-rs4926046	Overlapping clusters	VAAST	<i>GPRIN2</i>
COL23A1	Overlapping clusters	VAAST	<i>COL23A1</i>
rs147245242	Less than 90% success	VAAST, LD VAAST with rs66961966	<i>SLC38A10</i>

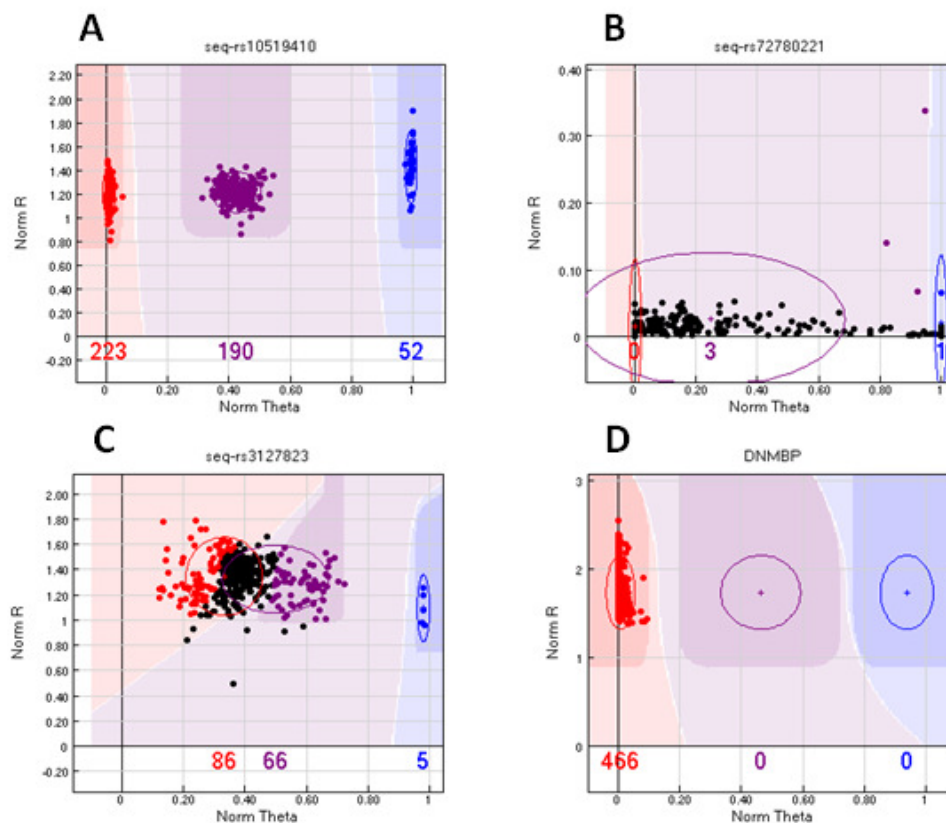
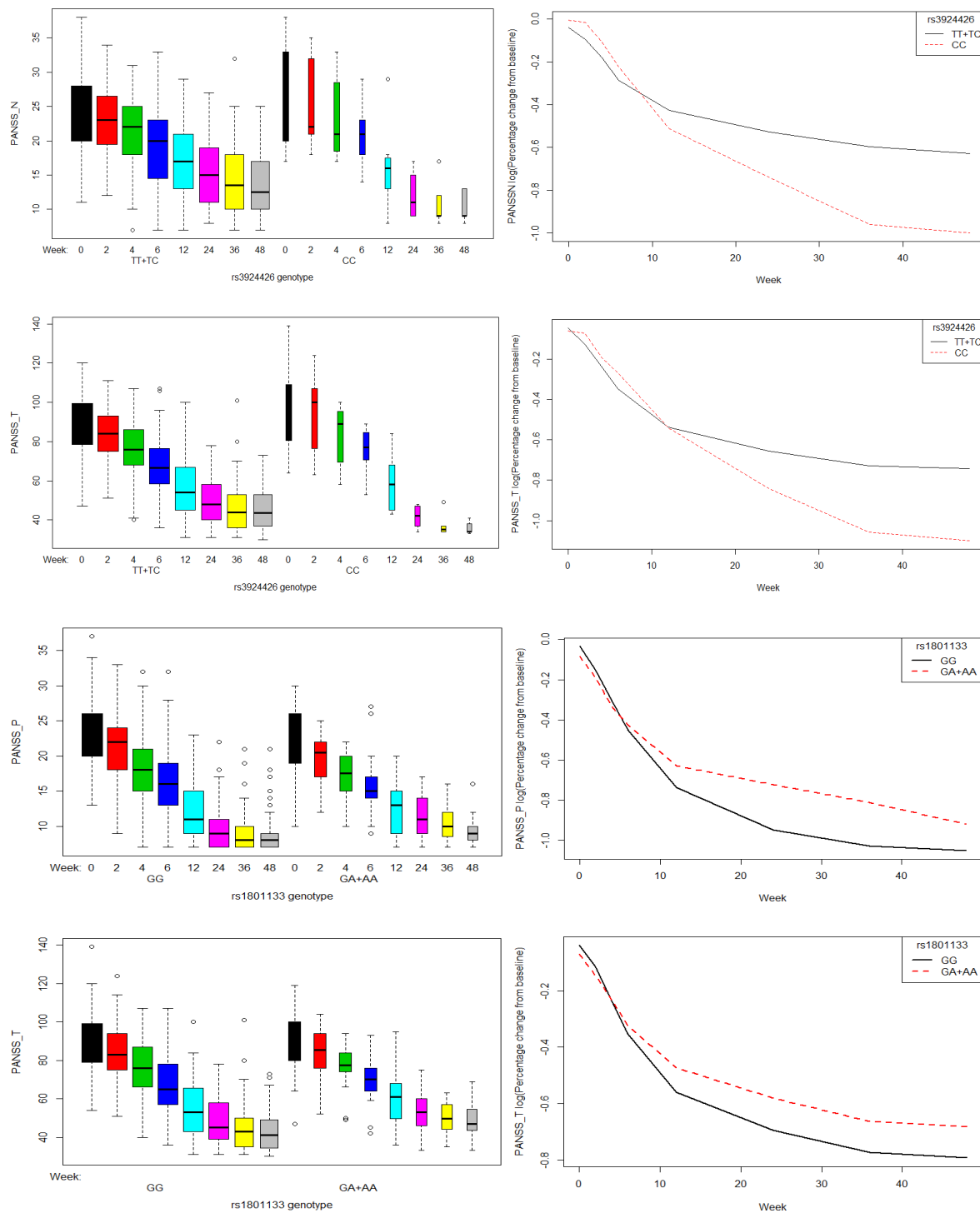


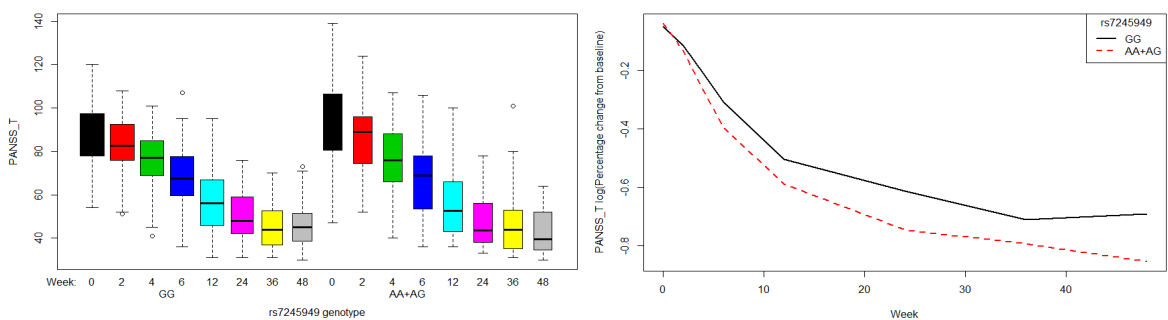
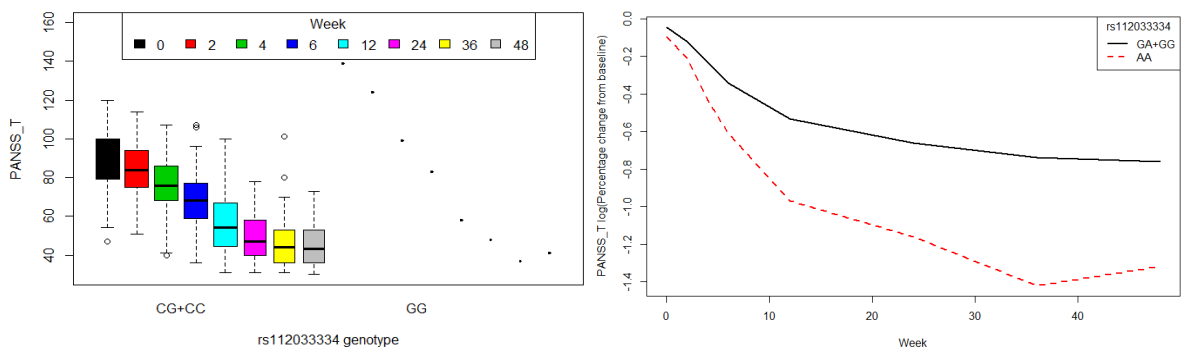
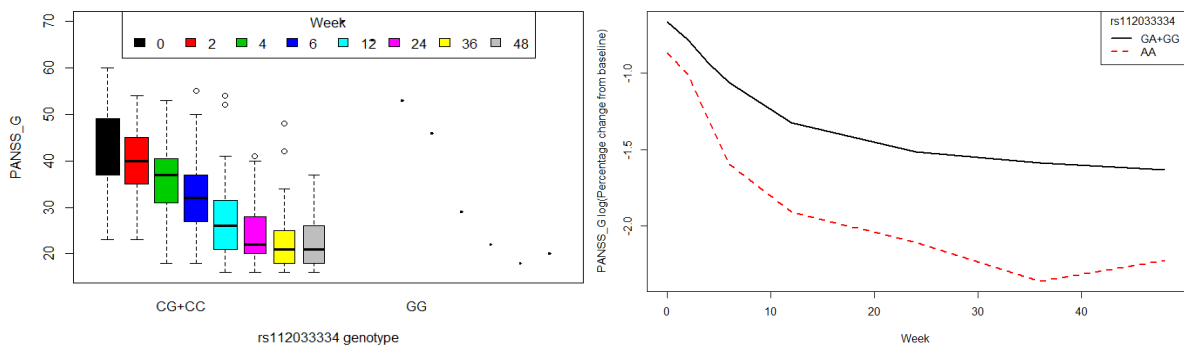
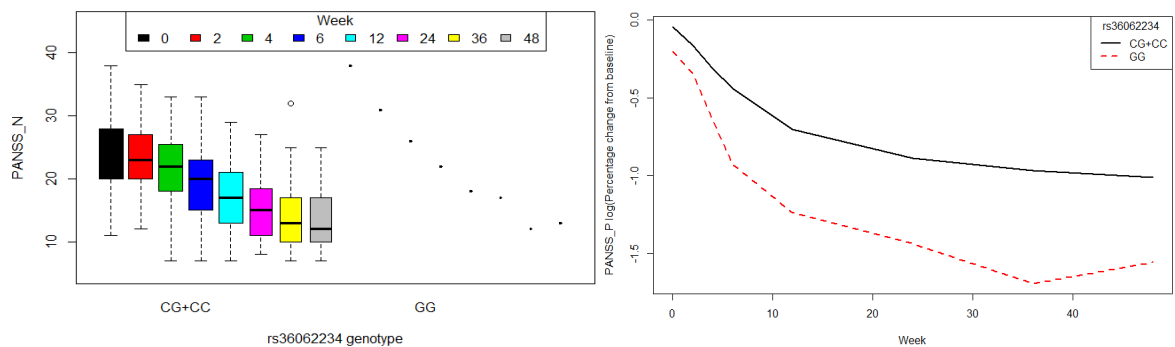
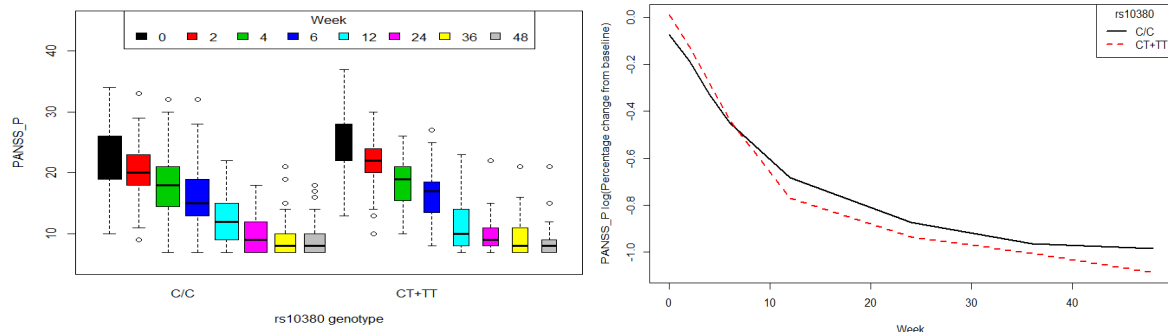
Figure S 7: BeadXpress genotyping clusters visualised in GenomeStudio. **A:** A successfully genotyped variant. **B:** A variant that failed genotyping due to a less than 90% success rate. **C:** A variant that failed genotyping due to overlapping clusters. **D:** A variant that could not be included in the analyses due to its absence in the cohorts of interest.

APPENDIX 10: Statistical analyses



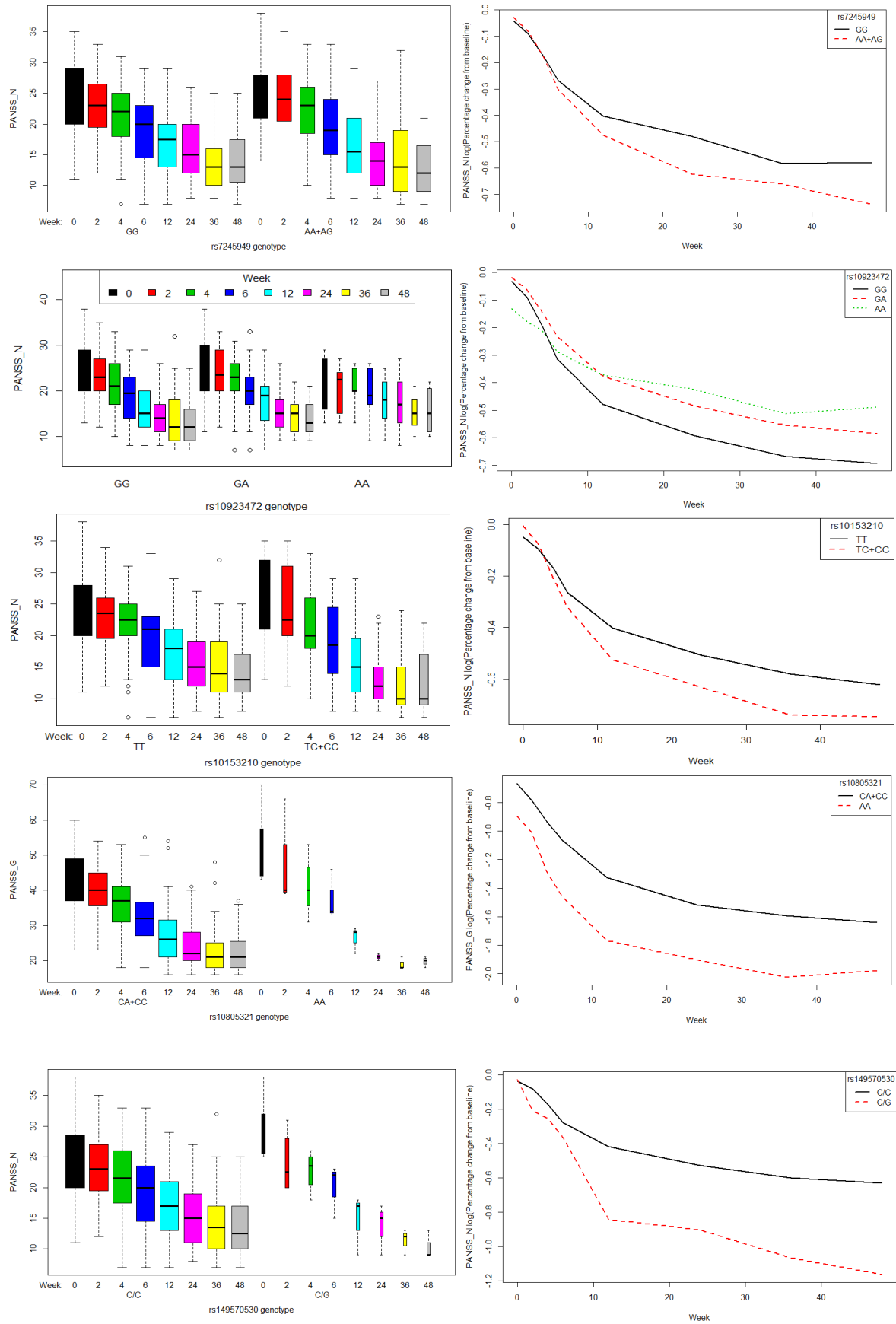
APPENDIX 10

STATISTICAL ANALYSES



APPENDIX 10

STATISTICAL ANALYSES



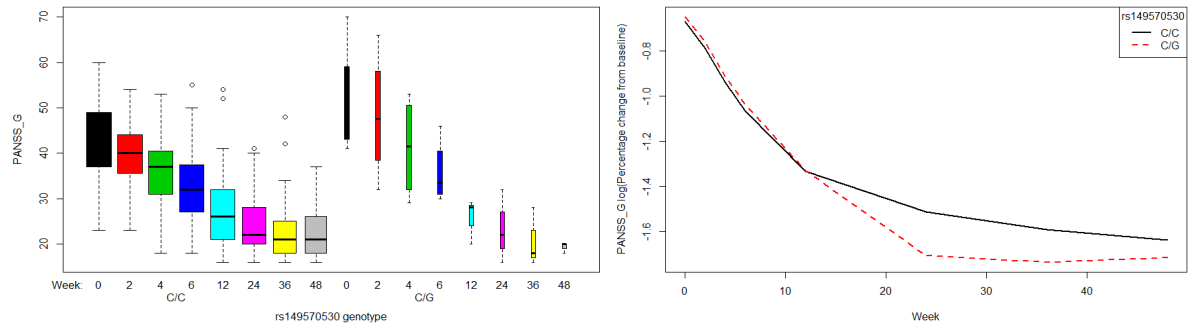


Figure S 8: Observed trait values represented as box plots (left) and corresponding percentage change from baseline in trait values (right), observed per week for each of the significantly associated variants. In the case of the box plots, the distribution of the PANSS scores recorded at each visit is displayed for each genotype class and the width of the box plots correlates to the number of individuals observed in each of these classes.

References

1. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL. Measuring the global burden of disease and risk factors, 1990–2001. In: *Global burden of disease and risk factors*. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ (Eds.). World Bank, Washington, DC (2006).
2. Saxena S, Skeen S. No health without mental health: challenges and opportunities in global mental health. *Afr J Psychiatry (Johannesbg)*. 15(6), 397–400 (2012).
3. Demyttenaere K, Bruffaerts R, Posada-Villa J, *et al*. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA*. 291(21), 2581–2590 (2004).
4. Lund C, Kleintjes S, Kakuma R, Flisher AJ. Public sector mental health systems in South Africa: inter-provincial comparisons and policy implications. *Soc Psychiatry Psychiatr Epidemiol*. 45(3), 393–404 (2010).
5. Global Forum for Health Research » 10/90 gap [Internet]. Available from: <http://www.globalforumhealth.org/about/1090-gap/>.
6. De Jesus Mari J, Patel V, Kieling C, *et al*. 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. *Afr J Psychiatry (Johannesbg)*. 12(1), 33–39 (2009).
7. Lund C, Tomlinson M, De Silva M, *et al*. PRIME: a programme to reduce the treatment gap for mental disorders in five low- and middle-income countries. *PLoS Med*. 9(12), e1001359 (2012).
8. Drew N, Faydi E, Freeman M, Kettaneh A, Ommeren MV. Mental health and development: Targeting people with mental health conditions as a vulnerable group [Internet]. (2010). Available from: http://www.who.int/mental_health/policy/mhtargeting/en/.
9. Rössler W, Salize HJ, van Os J, Riecher-Rössler A. Size of burden of schizophrenia and psychotic disorders. *Eur Neuropsychopharmacol*. 15(4), 399–409 (2005).
10. Tandon R, Keshavan MS, Nasrallah HA. Schizophrenia, “Just the Facts”: what we know in 2008 part 1: overview. *Schizophr. Res*. 100(1-3), 4–19 (2008).
11. Knapp M, Mangalore R, Simon J. The global costs of schizophrenia. *Schizophr Bull*. 30(2), 279–293 (2004).
12. Collins PY, Patel V, Joestl SS, *et al*. Grand challenges in global mental health. *Nature*. 475(7354), 27–30 (2011).
13. WHO | The global burden of disease: 2004 update [Internet]. WHO. Available from: http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/index.html.
14. WHO | mhGAP Mental Health Gap Action Programme [Internet]. Available from: http://www.who.int/mental_health/evidence/mhGAP/en/index.html.
15. Tandon R, Nasrallah HA, Keshavan MS. Schizophrenia, “just the facts” 5. Treatment and prevention. Past, present, and future. *Schizophr. Res*. 122(1-3), 1–23 (2010).
16. Arranz MJ, Munro JC. Toward understanding genetic risk for differential antipsychotic response in individuals with schizophrenia. *Expert Rev Clin Pharmacol*. 4(3), 389–405 (2011).
17. León-Cachón RBR, Ascacio-Martínez JA, Barrera-Saldaña HA. Individual response to drug therapy: bases and study approaches. *Rev. Invest. Clin*. 64(4), 364–376 (2012).
18. Closing the R&D gap in African health care - McKinsey Quarterly - Health Care - Pharmaceuticals [Internet]. Available from: http://www.mckinseyquarterly.com/Closing_the_RD_gap_in_African_health_care_2593.
19. Deverka PA, Vernon J, McLeod HL. Economic opportunities and challenges for pharmacogenomics. *Annu. Rev. Pharmacol. Toxicol*. 50, 423–437 (2010).
20. Roederer MW, Sanchez-Giron F, Kalideen K, Kudzi W, McLeod HL, Zhang W. Pharmacogenetics and rational drug use around the world. *Pharmacogenomics*. 12(6), 897–905 (2011).
21. The Pharmacogenomics Knowledge Base [PharmGKB] [Internet]. Available from: <http://www.pharmgkb.org/>.

REFERENCES

22. Genomics > Table of Pharmacogenomic Biomarkers in Drug Labels [Internet]. Available from: <http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm>.
23. Warfarin [PharmGKB] [Internet]. Available from: <http://www.pharmgkb.org/drug/PA451906>.
24. International Warfarin Pharmacogenetics Consortium, Klein TE, Altman RB, *et al.* Estimation of the warfarin dose with clinical and pharmacogenetic data. *N. Engl. J. Med.* 360(8), 753–764 (2009).
25. Illustration: Quantitative Advances Since the Human Genome Project (HGP) [Internet]. Available from: http://www.genome.gov/images/illustrations/hgp_measures.pdf.
26. Urban TJ. Whole-genome sequencing in pharmacogenetics. *Pharmacogenomics.* 14(4), 345–348 (2013).
27. Ritchie MD. The success of pharmacogenomics in moving genetic association studies from bench to bedside: study design and implementation of precision medicine in the post-GWAS era. *Hum. Genet.* 131(10), 1615–1626 (2012).
28. Ashley EA, Butte AJ, Wheeler MT, *et al.* Clinical assessment incorporating a personal genome. *Lancet.* 375(9725), 1525–1535 (2010).
29. ASHG 2012 Annual Meeting [Internet]. Available from: <http://www.ashg.org/2012meeting/>.
30. Altman RB. Pharmacogenomics: “noninferiority” is sufficient for initial implementation. *Clin. Pharmacol. Ther.* 89(3), 348–350 (2011).
31. WHO | Schizophrenia [Internet]. Available from: http://www.who.int/mental_health/management/schizophrenia/en/.
32. Brain Facts: A Primer on the Brain and Nervous System [Internet]. Available from: <http://www.brainline.org/content/2009/06/brain-facts-a-primer-on-the-brain-and-nervous-system.html>.
33. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *PLoS Med.* 2(5), e141 (2005).
34. Tandon R, Keshavan MS, Nasrallah HA. Schizophrenia, “just the facts” what we know in 2008. 2. Epidemiology and etiology. *Schizophr. Res.* 102(1-3), 1–18 (2008).
35. Owens DGC, Miller P, Lawrie SM, Johnstone EC. Pathogenesis of schizophrenia: a psychopathological perspective. *Br J Psychiatry.* 186, 386–393 (2005).
36. Tandon R, Nasrallah HA, Keshavan MS. Schizophrenia, “just the facts” 4. Clinical features and conceptualization. *Schizophr. Res.* 110(1-3), 1–23 (2009).
37. Phillips LJ, McGorry PD, Garner B, *et al.* Stress, the hippocampus and the hypothalamic-pituitary-adrenal axis: implications for the development of psychotic disorders. *Aust N Z J Psychiatry.* 40(9), 725–741 (2006).
38. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull.* 13(2), 261–276 (1987).
39. Lyne JP, Kinsella A, O’Donoghue B. Can we combine symptom scales for collaborative research projects? *J Psychiatr Res.* 46(2), 233–238 (2012).
40. McGorry PD, Hickie IB, Yung AR, Pantelis C, Jackson HJ. Clinical staging of psychiatric disorders: a heuristic framework for choosing earlier, safer and more effective interventions. *Aust N Z J Psychiatry.* 40(8), 616–622 (2006).
41. Tandon R. The nosology of schizophrenia: toward DSM-5 and ICD-11. *Psychiatr. Clin. North Am.* 35(3), 557–569 (2012).
42. Bondy B. Genetics in psychiatry: are the promises met? *World J. Biol. Psychiatry.* 12(2), 81–88 (2011).
43. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-IV-TR). American Psychiatric Association., Washington, DC (1994).
44. WHO | International Classification of Diseases (ICD) [Internet]. WHO. Available from: <http://www.who.int/classifications/icd/en/>.

REFERENCES

45. Silveira C, Marques-Teixeira J, de Bastos-Leite AJ. More than one century of schizophrenia: an evolving perspective. *J. Nerv. Ment. Dis.* 200(12), 1054–1057 (2012).
46. Van Os J, Kapur S. Schizophrenia. *Lancet.* 374(9690), 635–645 (2009).
47. Adam D. Mental health: On the spectrum. *Nature.* 496(7446), 416–418 (2013).
48. Clarke MC, Kelleher I, Clancy M, Cannon M. Predicting risk and the emergence of schizophrenia. *Psychiatr. Clin. North Am.* 35(3), 585–612 (2012).
49. Lund C, De Silva M, Plagerson S, *et al.* Poverty and mental disorders: breaking the cycle in low-income and middle-income countries. *Lancet.* 378(9801), 1502–1514 (2011).
50. Roos JL. Genetics of schizophrenia: communicating scientific findings in the clinical setting. *Afr J Psychiatry (Johannesbg).* 14(2), 105–111 (2011).
51. Lichtenstein P, Yip BH, Björk C, *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet.* 373(9659), 234–239 (2009).
52. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.* 13(8), 537–551 (2012).
53. SZGene [Internet]. Available from: <http://www.szgene.org/>.
54. Girard SL, Dion PA, Rouleau GA. Schizophrenia genetics: putting all the pieces together. *Curr Neurol Neurosci Rep.* 12(3), 261–266 (2012).
55. Gejman PV, Sanders AR, Kendler KS. Genetics of schizophrenia: new findings and challenges. *Annu Rev Genomics Hum Genet.* 12, 121–144 (2011).
56. Emsley R. Drugs in development for the treatment of schizophrenia. *Expert Opin Investig Drugs.* 18(8), 1103–1118 (2009).
57. Burmeister M, McInnis MG, Zöllner S. Psychiatric genetics: progress amid controversy. *Nat. Rev. Genet.* 9(7), 527–540 (2008).
58. Adeyemo A, Rotimi C. Genetic variants associated with complex human diseases show wide variation across multiple populations. *Public Health Genomics.* 13(2), 72–79 (2010).
59. Mowry BJ, Gratten J. The emerging spectrum of allelic variation in schizophrenia: current evidence and strategies for the identification and functional characterization of common and rare variants. *Mol. Psychiatry.* 18(1), 38–52 (2013).
60. PGC - Psychiatric Genomics Consortium [Internet]. Available from: <https://pgc.unc.edu/>.
61. Ripke S, O'Dushlaine C, Chambert K, *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet.* doi: 10.1038/ng.2742 (2013).
62. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* 43(10), 969–976 (2011).
63. Purcell SM, Wray NR, Stone JL, *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 460(7256), 748–752 (2009).
64. Ng SB, Turner EH, Robertson PD, *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature.* 461(7261), 272–276 (2009).
65. Xu B, Roos JL, Dexheimer P, *et al.* Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat. Genet.* 43(9), 864–868 (2011).
66. Xu B, Ionita-Laza I, Roos JL, *et al.* De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.* 44(12), 1365–1369 (2012).
67. Need AC, McEvoy JP, Gennarelli M, *et al.* Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *Am. J. Hum. Genet.* 91(2), 303–312 (2012).
68. Girard SL, Gauthier J, Noreau A, *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat. Genet.* 43(9), 860–863 (2011).
69. Gulsuner S, Walsh T, Watts AC, *et al.* Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell.* 154(3), 518–529 (2013).

REFERENCES

70. Timms AE, Dorschner MO, Wechsler J, *et al.* Support for the N -methyl-d-aspartate receptor hypofunction hypothesis of schizophrenia from exome sequencing in multiplex families. *JAMA Psychiatry.* 70(6), 582–590 (2013).
71. Abecasis GR, Auton A, Brooks LD, *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature.* 491(7422), 56–65 (2012).
72. Crespi B, Summers K, Dorus S. Adaptive evolution of genes underlying schizophrenia. *Proc. Biol. Sci.* 274(1627), 2801–2810 (2007).
73. Cacabelos R, Hashimoto R, Takeda M. Pharmacogenomics of antipsychotics efficacy for schizophrenia. *Psychiatry Clin. Neurosci.* 65(1), 3–19 (2011).
74. Carpenter WT, Koenig JI. The evolution of drug development in schizophrenia: past issues and future opportunities. *Neuropsychopharmacology.* 33(9), 2061–2079 (2008).
75. Tsuang MT, Woolson RF, Fleming JA. Long-term outcome of major psychoses. I. Schizophrenia and affective disorders compared with psychiatrically symptom-free surgical conditions. *Arch. Gen. Psychiatry.* 36(12), 1295–1301 (1979).
76. Hegarty JD, Baldessarini RJ, Tohen M, Waternaux C, Oepen G. One hundred years of schizophrenia: a meta-analysis of the outcome literature. *Am J Psychiatry.* 151(10), 1409–1416 (1994).
77. Cole J. Evaluation of drug treatments in psychiatry. *Proc Annu Meet Am Psychopathol Assoc.* 52, 14–31 (1964).
78. Leucht S, Barnes TRE, Kissling W, Engel RR, Correll C, Kane JM. Relapse prevention in schizophrenia with new-generation antipsychotics: a systematic review and exploratory meta-analysis of randomized, controlled trials. *Am J Psychiatry.* 160(7), 1209–1222 (2003).
79. Adams CE, Fenton MK, Quraishi S, David AS. Systematic meta-review of depot antipsychotic drugs for people with schizophrenia. *Br J Psychiatry.* 179, 290–299 (2001).
80. WHO | WHO Model Lists of Essential Medicines [Internet]. Available from: <http://www.who.int/medicines/publications/essentialmedicines/en/index.html>.
81. Laing R, Waning B, Gray A, Ford N, 't Hoen E. 25 years of the WHO essential medicines lists: progress and challenges. *Lancet.* 361(9370), 1723–1729 (2003).
82. Antipsychotic - PubMed - NCBI [Internet]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed>.
83. Correll CU. What are we looking for in new antipsychotics? *J Clin Psychiatry.* 72 Suppl 1, 9–13 (2011).
84. Kahn RS, Fleischhacker WW, Boter H, *et al.* Effectiveness of antipsychotic drugs in first-episode schizophrenia and schizophreniform disorder: an open randomised clinical trial. *Lancet.* 371(9618), 1085–1097 (2008).
85. Mas S, Llerena A, Saíz J, Bernardo M, Lafuente A. Strengths and weaknesses of pharmacogenetic studies of antipsychotic drugs: the potential value of the PEPs study. *Pharmacogenomics.* 13(15), 1773–1782 (2012).
86. The International Psychopharmacology Algorithm Project (IPAP) [Internet]. Available from: <http://www.ipap.org/index.php>.
87. Kennedy JL, Voudouris NC. Incorporating psychiatric pharmacogenetics into family practice. *Pharmacogenomics.* 14(10), 1121–1124 (2013).
88. Buckley PF. Factors that influence treatment success in schizophrenia. *J Clin Psychiatry.* 69 Suppl 3, 4–10 (2008).
89. Liu-Seifert H, Adams DH, Kinon BJ. Discontinuation of treatment of schizophrenic patients is driven by poor symptom response: a pooled post-hoc analysis of four atypical antipsychotic drugs. *BMC Med.* 3, 21 (2005).
90. Crilly J. The history of clozapine and its emergence in the US market: a review and analysis. *Hist Psychiatry.* 18(1), 39–60 (2007).
91. Kane J, Honigfeld G, Singer J, Meltzer H. Clozapine for the treatment-resistant schizophrenic. A double-blind comparison with chlorpromazine. *Arch. Gen. Psychiatry.* 45(9), 789–796 (1988).

REFERENCES

92. Meltzer HY, Alphas L, Green AI, *et al.* Clozapine treatment for suicidality in schizophrenia: International Suicide Prevention Trial (InterSePT). *Arch. Gen. Psychiatry.* 60(1), 82–91 (2003).
93. Taylor M, Cavanagh J, Hodgson R, Tiihonen J. Examining the effectiveness of antipsychotic medication in first-episode psychosis. *J. Psychopharmacol. (Oxford).* 26(5 Suppl), 27–32 (2012).
94. Leucht S, Corves C, Arbter D, Engel RR, Li C, Davis JM. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. *Lancet.* 373(9657), 31–41 (2009).
95. Agius M, Davis A, Gilhooley M, Chapman S, Zaman R. What do large scale studies of medication in schizophrenia add to our management strategies? *Psychiatr Danub.* 22(2), 323–328 (2010).
96. Correll CU. From receptor pharmacology to improved outcomes: individualising the selection, dosing, and switching of antipsychotics. *Eur. Psychiatry.* 25 Suppl 2, S12–21 (2010).
97. Kapur S, Zipursky R, Jones C, Remington G, Houle S. Relationship between dopamine D(2) occupancy, clinical response, and side effects: a double-blind PET study of first-episode schizophrenia. *Am J Psychiatry.* 157(4), 514–520 (2000).
98. Tandon R. Antipsychotics in the treatment of schizophrenia: an overview. *J Clin Psychiatry.* 72 Suppl 1, 4–8 (2011).
99. Brenner HD, Dencker SJ, Goldstein MJ, *et al.* Defining treatment refractoriness in schizophrenia. *Schizophr Bull.* 16(4), 551–561 (1990).
100. Reynolds GP. The pharmacogenetics of symptom response to antipsychotic drugs. *Psychiatry Investig.* 9(1), 1–7 (2012).
101. Robinson D, Woerner MG, Alvir JM, *et al.* Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. *Arch. Gen. Psychiatry.* 56(3), 241–247 (1999).
102. Suzuki T, Remington G, Mulsant BH, *et al.* Defining treatment-resistant schizophrenia and response to antipsychotics: a review and recommendation. *Psychiatry Res.* 197(1-2), 1–6 (2012).
103. Case M, Stauffer VL, Ascher-Svanum H, *et al.* The heterogeneity of antipsychotic response in the treatment of schizophrenia. *Psychol Med.* 41(6), 1291–1300 (2011).
104. Andreasen NC, Carpenter WT Jr, Kane JM, Lasser RA, Marder SR, Weinberger DR. Remission in schizophrenia: proposed criteria and rationale for consensus. *Am J Psychiatry.* 162(3), 441–449 (2005).
105. Csernansky JG, Mahmoud R, Brenner R. A comparison of risperidone and haloperidol for the prevention of relapse in patients with schizophrenia. *N. Engl. J. Med.* 346(1), 16–22 (2002).
106. Albus M. Clinical courses of schizophrenia. *Pharmacopsychiatry.* 45 Suppl 1, S31–35 (2012).
107. Emsley R, Oosthuizen P, Koen L, Niehaus D, Martinez L. Comparison of treatment response in second-episode versus first-episode schizophrenia. *J Clin Psychopharmacol.* 33(1), 80–83 (2013).
108. Weiden PJ, Olfson M. Cost of relapse in schizophrenia. *Schizophr Bull.* 21(3), 419–429 (1995).
109. Ascher-Svanum H, Zhu B, Faries DE, *et al.* The cost of relapse and the predictors of relapse in the treatment of schizophrenia. *BMC Psychiatry.* 10, 2 (2010).
110. Emsley R. Early intervention in the management of schizophrenia: introduction. *Early Interv Psychiatry.* 3 Suppl 1, S1–2 (2009).
111. Malhi G, Adams D, Moss B, Walter G. To medicate or not to medicate, when diagnosis is in question: decision-making in first episode psychosis. *Australas Psychiatry.* 18(3), 230–237 (2010).
112. Olivares JM, Alptekin K, Azorin J-M, *et al.* Psychiatrists' awareness of adherence to antipsychotic medication in patients with schizophrenia: results from a survey conducted across Europe, the Middle East, and Africa. *Patient Prefer Adherence.* 7, 121–132 (2013).
113. Robinson D. First-episode schizophrenia. *CNS Spectr.* 15(4 Suppl 6), 4–7 (2010).
114. Emsley R. New advances in pharmacotherapy for early psychosis. *Early Interv Psychiatry.* 3 Suppl 1, S8–12 (2009).

REFERENCES

115. Emsley R, Chiliza B, Asmal L, Mashile M, Fusar-Poli P. Long-acting injectable antipsychotics in early psychosis: a literature review. *Early Interv Psychiatry*. 7(3), 247–254 (2013).
116. Liu C-C, Chiang H-L, Wu S-Y. Applying long-acting injectable antipsychotics in first-episode schizophrenia to achieve first remission. *CNS Drugs*. 22(3), 257–258 (2008).
117. Emsley R, Chiliza B, Schoeman R. Predictors of long-term outcome in schizophrenia. *Curr Opin Psychiatry*. 21(2), 173–177 (2008).
118. Emsley R, Oosthuizen P, Koen L, Niehaus DJH, Medori R, Rabinowitz J. Remission in patients with first-episode schizophrenia receiving assured antipsychotic medication: a study with risperidone long-acting injection. *Int Clin Psychopharmacol*. 23(6), 325–331 (2008).
119. Zhang J-P, Malhotra AK. Pharmacogenetics and antipsychotics: therapeutic efficacy and side effects prediction. *Expert Opin Drug Metab Toxicol*. 7(1), 9–37 (2011).
120. Gesteira A, Barros F, Martín A, *et al*. Pharmacogenetic studies on the antipsychotic treatment. Current status and perspectives. *Actas Esp Psiquiatr*. 38(5), 301–316 (2010).
121. Chowdhury NI, Remington G, Kennedy JL. Genetics of antipsychotic-induced side effects and agranulocytosis. *Curr Psychiatry Rep*. 13(2), 156–165 (2011).
122. Nordström AL, Farde L, Wiesel FA, *et al*. Central D2-dopamine receptor occupancy in relation to antipsychotic drug effects: a double-blind PET study of schizophrenic patients. *Biol Psychiatry*. 33(4), 227–235 (1993).
123. Casey DE. Tardive dyskinesia. *West. J. Med*. 153(5), 535–541 (1990).
124. Schooler NR, Kane JM. Research diagnoses for tardive dyskinesia. *Arch. Gen. Psychiatry*. 39(4), 486–487 (1982).
125. Holloman LC, Marder SR. Management of acute extrapyramidal effects induced by antipsychotic drugs. *Am J Health Syst Pharm*. 54(21), 2461–2477 (1997).
126. Hummer M, Huber J. Hyperprolactinaemia and antipsychotic therapy in schizophrenia. *Curr Med Res Opin*. 20(2), 189–197 (2004).
127. Balt SL, Galloway GP, Baggott MJ, Schwartz Z, Mendelson J. Mechanisms and genetics of antipsychotic-associated weight gain. *Clin. Pharmacol. Ther*. 90(1), 179–183 (2011).
128. Ryan MCM, Flanagan S, Kinsella U, Keeling F, Thakore JH. The effects of atypical antipsychotics on visceral fat distribution in first episode, drug-naive patients with schizophrenia. *Life Sci*. 74(16), 1999–2008 (2004).
129. Reist C, Mintz J, Albers LJ, Jamal MM, Szabo S, Ozdemir V. Second-generation antipsychotic exposure and metabolic-related disorders in patients with schizophrenia: an observational pharmacoepidemiology study from 1988 to 2002. *J Clin Psychopharmacol*. 27(1), 46–51 (2007).
130. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999–2010. *JAMA*. 307(5), 491–497 (2012).
131. De Hert M, Dekker JM, Wood D, Kahl KG, Holt RIG, Möller H-J. Cardiovascular disease and diabetes in people with severe mental illness position statement from the European Psychiatric Association (EPA), supported by the European Association for the Study of Diabetes (EASD) and the European Society of Cardiology (ESC). *Eur. Psychiatry*. 24(6), 412–424 (2009).
132. Lin PI, Shuldiner AR. Rethinking the genetic basis for comorbidity of schizophrenia and type 2 diabetes. *Schizophr. Res*. 123(2-3), 234–243 (2010).
133. American Diabetes Association, American Psychiatric Association, American Association of Clinical Endocrinologists, North American Association for the Study of Obesity. Consensus development conference on antipsychotic drugs and obesity and diabetes. *Diabetes Care*. 27(2), 596–601 (2004).
134. Correll CU, Lencz T, Malhotra AK. Antipsychotic drugs and obesity. *Trends Mol Med*. 17(2), 97–107 (2011).
135. Alvir JM, Lieberman JA, Safferman AZ, Schwimmer JL, Schaaf JA. Clozapine-induced agranulocytosis. Incidence and risk factors in the United States. *N. Engl. J. Med*. 329(3), 162–167 (1993).

REFERENCES

136. Opgen-Rhein C, Dettling M. Clozapine-induced agranulocytosis and its genetic determinants. *Pharmacogenomics*. 9(8), 1101–1111 (2008).
137. Idänpään-Heikkilä J, Alhava E, Olkinuora M, Palva IP. Agranulocytosis during treatment with chlozapine. *Eur. J. Clin. Pharmacol.* 11(3), 193–198 (1977).
138. Chandrasekaran PK. Agranulocytosis monitoring with Clozapine patients: to follow guidelines or to attempt therapeutic controversies? *Singapore Med J.* 49(2), 96–99 (2008).
139. Vogel F. Moderne Probleme der Humangenetik. In: *Ergebnisse der Inneren Medizin und Kinderheilkunde*. Heilmeyer L, Schoen R, Rudder B de (Eds.). Springer Berlin Heidelberg, 52–125 (1959).
140. De Leon J. The future (or lack of future) of personalized prescription in psychiatry. *Pharmacol. Res.* 59(2), 81–89 (2009).
141. Pirmohamed M. Pharmacogenetics and pharmacogenomics. *Br J Clin Pharmacol.* 52(4), 345–347 (2001).
142. Vojvoda D, Grimmell K, Sernyak M, Mazure CM. Monozygotic twins concordant for response to clozapine. *Lancet.* 347(8993), 61 (1996).
143. Horáček JL. Clozapine-induced concordant agranulocytosis in monozygotic twins. *Int J Psychiatry Clin Pract.* 5(1), 71–73 (2001).
144. Wehmeier PM, Gebhardt S, Schmidtke J, Remschmidt H, Hebebrand J, Theisen FM. Clozapine: weight gain in a pair of monozygotic twins concordant for schizophrenia and mild mental retardation. *Psychiatry Res.* 133(2-3), 273–276 (2005).
145. Waddington JL, Youssef HA. The expression of schizophrenia, affective disorder and vulnerability to tardive dyskinesia in an extensive pedigree. *Br J Psychiatry.* 153, 376–381 (1988).
146. Müller DJ, Schulze TG, Knapp M, *et al.* Familial occurrence of tardive dyskinesia. *Acta Psychiatr Scand.* 104(5), 375–379 (2001).
147. Franchini L, Serretti A, Gasperini M, Smeraldi E. Familial concordance of fluvoxamine response as a tool for differentiating mood disorder pedigrees. *J Psychiatr Res.* 32(5), 255–259 (1998).
148. Malhotra AK, Lencz T, Correll CU, Kane JM. Genomics and the future of pharmacotherapy in psychiatry. *Int Rev Psychiatry.* 19(5), 523–530 (2007).
149. De Leon J, Susce MT, Johnson M, *et al.* DNA microarray technology in the clinical environment: the AmpliChip CYP450 test for CYP2D6 and CYP2C19 genotyping. *CNS Spectr.* 14(1), 19–34 (2009).
150. Arif SA, Mitchell MM. Iloperidone: A new drug for the treatment of schizophrenia. *Am J Health Syst Pharm.* 68(4), 301–308 (2011).
151. Antipsychotics [PharmGKB] [Internet]. Available from: <http://www.pharmgkb.org/drug/PA452233>.
152. Genomas, Inc. [Internet]. Available from: <http://www.genomas.net/phyziotype-pims.php>.
153. Genomas, Inc. [Internet]. Available from: <http://www.genomas.net/phyziotype-hilomet.php>.
154. Clinical Data Launches Pharmacogenetic Test for Clozapine-Induced Agranulocytosis on Schedule [Internet]. Available from: <http://www.thefreelibrary.com/Clinical+Data+Launches+Pharmacogenetic+Test+for+Clozapine+Induced...-a0158151979>.
155. LGC s Clozapine Efficacy Dx Could Support Clinical Data s Biomarker Research of Drug | Pharmacogenomics Reporter | Clinical Genomics | GenomeWeb [Internet]. Available from: <http://www.genomeweb.com/dxpgx/lgc-s-clozapine-efficacy-dx-could-support-clinical-data-s-biomarker-research-dru>.
156. Luminex [Internet]. Available from: http://www.luminexcorp.com/product_list.html?record_id=335&product_type=analyte.
157. Roche - AmpliChip® CYP450 Test [Internet]. Available from: <http://www.roche.com/products/product-details.htm?type=product&id=17>.
158. GeneSightRx test [Internet]. Available from: <http://www.assureroxhealth.com/products>.

159. Clinicians [Internet]. Available from: <http://www.suregenetest.com/Clinicians/Clinicians.aspx>.
160. Progenika - PHARMAchip [Internet]. Available from: http://www.progenika.com/eu/index.php?option=com_content&task=view&id=144&Itemid=188.
161. Genetic Analysis, ADME PGx Genotyping, Somatic Mutation Profiling, SNP Genotyping, DNA Methylation Analysis, Molecular Typing, Gene Expression Analysis, GQE, MALDI-TOF, Mass Spectrometry | Sequenom, Inc. [Internet]. Available from: <http://www.sequenom.com/ADME-PGx-Genotyping>.
162. VeraCode ADME Core Panel - datasheet_veracode_adme_core_panel.pdf [Internet]. Available from: http://www.illumina.com/Documents/products/datasheets/datasheet_veracode_adme_core_panel.pdf.
163. Neuropharmagen [Internet]. Available from: <http://international.neurofarmagen.com//index.php?id=11&idioma=EN>.
164. CAMH: In a Canadian first, family practice offers genetic testing with CAMH to predict which psychiatric meds work best [Internet]. Available from: http://www.camh.ca/en/hospital/about_camh/newsroom/news_releases_media_advisories_and_backgrounders/current_year/Pages/In-a-Canadian-first,-family-practice-offers-genetic-testing-with-CAMH-to-predict-which-psychiatric-meds-work-best-.aspx.
165. Mallal S, Phillips E, Carosi G, *et al.* HLA-B*5701 screening for hypersensitivity to abacavir. *N. Engl. J. Med.* 358(6), 568–579 (2008).
166. Saag M, Balu R, Phillips E, *et al.* High sensitivity of human leukocyte antigen-b*5701 as a marker for immunologically confirmed abacavir hypersensitivity in white and black patients. *Clin. Infect. Dis.* 46(7), 1111–1118 (2008).
167. Jürgens G, Jacobsen CB, Rasmussen HB, Werge T, Nordentoft M, Andersen SE. Utility and adoption of CYP2D6 and CYP2C19 genotyping and its translation into psychiatric clinical practice. *Acta Psychiatr Scand.* 125(3), 228–237 (2012).
168. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin. Pharmacol. Ther.* 89(3), 464–467 (2011).
169. Kirchheiner J, Nickchen K, Bauer M, *et al.* Pharmacogenetics of antidepressants and antipsychotics: the contribution of allelic variations to the phenotype of drug response. *Mol. Psychiatry.* 9(5), 442–473 (2004).
170. Browse VIP Genes [PharmGKB] [Internet]. Available from: <http://www.pharmgkb.org/search/browseVip.action?browseKey=vipGenes>.
171. Arranz MJ, de Leon J. Pharmacogenetics and pharmacogenomics of schizophrenia: a review of last decade of research. *Mol. Psychiatry.* 12(8), 707–747 (2007).
172. Arranz MJ, Kapur S. Pharmacogenetics in psychiatry: are we ready for widespread clinical use? *Schizophr Bull.* 34(6), 1130–1144 (2008).
173. CYP2D6 [Internet]. Available from: <http://www.cypalleles.ki.se/cyp2d6.htm>.
174. Sistonen J, Fuselli S, Palo JU, Chauhan N, Padh H, Sajantila A. Pharmacogenetic variation at CYP2C9, CYP2C19, and CYP2D6 at global and microgeographic scales. *Pharmacogenet. Genomics.* 19(2), 170–179 (2009).
175. Stingl JC, Brockmöller J, Viviani R. Genetic variability of drug-metabolizing enzymes: the dual impact on psychiatric therapy and regulation of brain function. *Mol. Psychiatry.* 18(3), 273–287 (2013).
176. Arranz MJ, Rivera M, Munro JC. Pharmacogenetics of response to antipsychotics in patients with schizophrenia. *CNS Drugs.* 25(11), 933–969 (2011).
177. Foster A, Miller DD, Buckley P. Pharmacogenetics and schizophrenia. *Clin. Lab. Med.* 30(4), 975–993 (2010).
178. Rege S. Antipsychotic induced weight gain in schizophrenia: mechanisms and management. *Aust N Z J Psychiatry.* 42(5), 369–381 (2008).

REFERENCES

179. Gerretsen P, Müller DJ, Tiwari A, Mamo D, Pollock BG. The intersection of pharmacology, imaging, and genetics in the development of personalized medicine. *Dialogues Clin Neurosci.* 11(4), 363–376 (2009).
180. Risselada AJ, Mulder H, Heerdink ER, Egberts TCG. Pharmacogenetic testing to predict antipsychotic-induced weight gain: a systematic review. *Pharmacogenomics.* 12(8), 1213–1227 (2011).
181. Lett TAP, Wallace TJM, Chowdhury NI, Tiwari AK, Kennedy JL, Müller DJ. Pharmacogenetics of antipsychotic-induced weight gain: review and clinical implications. *Mol. Psychiatry.* 17(3), 242–266 (2012).
182. Moons T, de Roo M, Claes S, Dom G. Relationship between P-glycoprotein and second-generation antipsychotics. *Pharmacogenomics.* 12(8), 1193–1211 (2011).
183. Lee H-J, Kang S-G. Genetics of tardive dyskinesia. *Int. Rev. Neurobiol.* 98, 231–264 (2011).
184. Plesnicar BK. Personalized antipsychotic treatment: the adverse effects perspectives. *Psychiatr Danub.* 22(2), 329–334 (2010).
185. Thelma B, Srivastava V, Tiwari AK. Genetic underpinnings of tardive dyskinesia: passing the baton to pharmacogenetics. *Pharmacogenomics.* 9(9), 1285–1306 (2008).
186. Lohoff FW, Ferraro TN. Pharmacogenetic considerations in the treatment of psychiatric disorders. *Expert Opin Pharmacother.* 11(3), 423–439 (2010).
187. Lencz T, Malhotra AK. Pharmacogenetics of antipsychotic-induced side effects. *Dialogues Clin Neurosci.* 11(4), 405–415 (2009).
188. Bishop JR, Bishop DL. Iloperidone for the treatment of schizophrenia. *Drugs Today.* 46(8), 567–579 (2010).
189. Reynolds GP. The impact of pharmacogenetics on the development and use of antipsychotic drugs. *Drug Discov. Today.* 12(21-22), 953–959 (2007).
190. Maier W, Zobel A. Contribution of allelic variations to the phenotype of response to antidepressants and antipsychotics. *Eur Arch Psychiatry Clin Neurosci.* 258 Suppl 1, 12–20 (2008).
191. Steimer W. Pharmacogenetics and psychoactive drug therapy: ready for the patient? *Ther Drug Monit.* 32(4), 381–386 (2010).
192. Dorado P, Peñas-Lledó EM, Llerena A. CYP2D6 polymorphism: implications for antipsychotic drug response, schizophrenia and personality traits. *Pharmacogenomics.* 8(11), 1597–1608 (2007).
193. Di Giorgio A, Sambataro F, Bertolino A. Functional imaging as a tool to investigate the relationship between genetic variation and response to treatment with antipsychotics. *Curr. Pharm. Des.* 15(22), 2560–2572 (2009).
194. Shiroma PR, Geda YE, Mrazek DA. Pharmacogenomic implications of variants of monoaminergic-related genes in geriatric psychiatry. *Pharmacogenomics.* 11(9), 1305–1330 (2010).
195. Blanc O, Brousse G, Meary A, Leboyer M, Llorca P-M. Pharmacogenetic of response efficacy to antipsychotics in schizophrenia: pharmacodynamic aspects. Review and implications for clinical research. *Fundam Clin Pharmacol.* 24(2), 139–160 (2010).
196. Burdick KE, Gopin CB, Malhotra AK. Pharmacogenetic approaches to cognitive enhancement in schizophrenia. *Harv Rev Psychiatry.* 19(3), 102–108 (2011).
197. Sagud M, Mück-Seler D, Mihaljević-Peles A, et al. Catechol-O-methyl transferase and schizophrenia. *Psychiatr Danub.* 22(2), 270–274 (2010).
198. Vyas NS, Shamsi SA, Malhotra AK, Aitchison KJ, Kumari V. Can genetics inform the management of cognitive deficits in schizophrenia? *J. Psychopharmacol. (Oxford).* 26(3), 334–348 (2012).
199. Rybakowski JK. BDNF gene: functional Val66Met polymorphism in mood disorders and schizophrenia. *Pharmacogenomics.* 9(11), 1589–1593 (2008).

REFERENCES

200. Chagnon YC, Mérette C, Bouchard RH, Emond C, Roy M-A, Maziade M. A genome wide linkage study of obesity as secondary effect of antipsychotics in multigenerational families of eastern Quebec affected by psychoses. *Mol. Psychiatry*. 9(12), 1067–1074 (2004).
201. Inada T, Koga M, Ishiguro H, *et al.* Pathway-based association analysis of genome-wide screening data suggest that genes associated with the gamma-aminobutyric acid receptor signaling pathway are involved in neuroleptic-induced, treatment-resistant tardive dyskinesia. *Pharmacogenet. Genomics*. 18(4), 317–323 (2008).
202. HuGENavigator|GWAS Integrator|Search [Internet]. Available from: <http://www.hugenavigator.net/HuGENavigator/gWAHit.do?query=antipsychotics&Mysubmit=simple&geneOrderType=geneA>.
203. Malhotra AK, Correll CU, Chowdhury NI, *et al.* Association between common variants near the melanocortin 4 receptor gene and severe antipsychotic drug-induced weight gain. *Arch. Gen. Psychiatry*. 69(9), 904–912 (2012).
204. Adkins DE, Aberg K, McClay JL, *et al.* Genomewide pharmacogenomic study of metabolic side effects to antipsychotic drugs. *Mol. Psychiatry*. 16(3), 321–332 (2011).
205. Athanasiu L, Brown AA, Birkenaes AB, *et al.* Genome-wide association study identifies genetic loci associated with body mass index and high density lipoprotein-cholesterol levels during psychopharmacological treatment - a cross-sectional naturalistic study. *Psychiatry Res*. 197(3), 327–336 (2012).
206. Aberg K, Adkins DE, Bukszár J, *et al.* Genomewide association study of movement-related adverse antipsychotic effects. *Biol. Psychiatry*. 67(3), 279–282 (2010).
207. McClay JL, Adkins DE, Aberg K, *et al.* Genome-wide pharmacogenomic study of neurocognition as an indicator of antipsychotic treatment response in schizophrenia. *Neuropsychopharmacology*. 36(3), 616–626 (2011).
208. McClay JL, Adkins DE, Aberg K, *et al.* Genome-wide pharmacogenomic analysis of response to treatment with antipsychotics. *Mol. Psychiatry*. 16(1), 76–85 (2011).
209. Aberg K, Adkins DE, Liu Y, *et al.* Genome-wide association study of antipsychotic-induced QTc interval prolongation. *Pharmacogenomics J*. 12(2), 165–172 (2012).
210. Volpi S, Heaton C, Mack K, *et al.* Whole genome association study identifies polymorphisms associated with QT prolongation during iloperidone treatment of schizophrenia. *Mol. Psychiatry*. 14(11), 1024–1031 (2009).
211. Alkelai A, Greenbaum L, Rigbi A, Kanyas K, Lerer B. Genome-wide association study of antipsychotic-induced parkinsonism severity among schizophrenia patients. *Psychopharmacology (Berl.)*. 206(3), 491–499 (2009).
212. Drago A, Crisafulli C, Serretti A. The genetics of antipsychotic induced tremors: a genome-wide pathway analysis on the STEP-BD SCP sample. *Am. J. Med. Genet. B Neuropsychiatr. Genet*. 156B(8), 975–986 (2011).
213. Greenbaum L, Alkelai A, Rigbi A, Kohn Y, Lerer B. Evidence for association of the GLI2 gene with tardive dyskinesia in patients with chronic schizophrenia. *Mov. Disord*. 25(16), 2809–2817 (2010).
214. Lavedan C, Licamele L, Volpi S, *et al.* Association of the NPAS3 gene and five other loci with response to the antipsychotic iloperidone identified in a whole genome association study. *Mol. Psychiatry*. 14(8), 804–819 (2009).
215. Daly AK. Genome-wide association studies in pharmacogenomics. *Nat. Rev. Genet*. 11(4), 241–246 (2010).
216. Samuels DC, Burn DJ, Chinnery PF. Detecting new neurodegenerative disease genes: does phenotype accuracy limit the horizon? *Trends Genet*. 25(11), 486–488 (2009).
217. Rietschel M, Kennedy JL, Macciardi F, Meltzer HY. Application of pharmacogenetics to psychotic disorders: the first consensus conference. The Consensus Group for Outcome Measures in Psychoses for Pharmacological Studies. *Schizophr. Res*. 37(2), 191–196 (1999).

REFERENCES

218. Zhang J-P, Malhotra AK. Pharmacogenetics of antipsychotics: recent progress and methodological issues. *Expert Opin Drug Metab Toxicol.* 9(2), 183–191 (2013).
219. Patel V, Kleinman A. Poverty and common mental disorders in developing countries. *Bull. World Health Organ.* 81(8), 609–615 (2003).
220. Stein DJ, Seedat S, Herman A, *et al.* Lifetime prevalence of psychiatric disorders in South Africa. *Br J Psychiatry.* 192(2), 112–117 (2008).
221. Burns J. Dispelling a myth: developing world poverty, inequality, violence and social fragmentation are not good for outcome in schizophrenia. *Afr J Psychiatry (Johannesbg).* 12(3), 200–205 (2009).
222. Kapp C. Crystal meth boom adds to South Africa's health challenges. *Lancet.* 371(9608), 193–194 (2008).
223. Emsley R. Focus on psychiatry in South Africa. *Br J Psychiatry.* 178, 382–386 (2001).
224. Seedat S, Williams DR, Herman AA, *et al.* Mental health service use among South Africans for mood, anxiety and substance use disorders. *S. Afr. Med. J.* 99(5 Pt 2), 346–352 (2009).
225. Sorsdahl KR, Stein DJ. Knowledge of and stigma associated with mental disorders in a South African community sample. *J. Nerv. Ment. Dis.* 198(10), 742–747 (2010).
226. Mbanga NI, Niehaus DJH, Mzamo NC, *et al.* Attitudes towards and beliefs about schizophrenia in Xhosa families with affected probands. *Curationis.* 25(1), 69–73 (2002).
227. Botha UA, Koen L, Niehaus DJH. Perceptions of a South African schizophrenia population with regards to community attitudes towards their illness. *Soc Psychiatry Psychiatr Epidemiol.* 41(8), 619–623 (2006).
228. Niehaus DJH, Oosthuizen P, Lochner C, *et al.* A culture-bound syndrome “amafufunyana” and a culture-specific event “ukuthwasa”: differentiated by a family history of schizophrenia and other psychiatric disorders. *Psychopathology.* 37(2), 59–63 (2004).
229. Mosotho L, Louw D, Calitz FJ. Schizophrenia among Sesotho speakers in South Africa. *Afr J Psychiatry (Johannesbg).* 14(1), 50–55 (2011).
230. Warnich L, Drögemöller BI, Pepper MS, Dandara C, Wright GEB. Pharmacogenomic research in South Africa: Lessons learned and future opportunities in the rainbow nation. *Curr Pharmacogenomics Person Med.* 9(3), 191–207 (2011).
231. Burns JK, Jhazbhay K, Kidd M, Emsley RA. Causal attributions, pathway to care and clinical features of first-episode psychosis: a South African perspective. *Int J Soc Psychiatry.* 57(5), 538–545 (2011).
232. Perkins DO, Gu H, Boteva K, Lieberman JA. Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. *Am J Psychiatry.* 162(10), 1785–1804 (2005).
233. Koen L, Magni P, Niehaus DJ, le Roux A. Antipsychotic prescription patterns in Xhosa patients with schizophrenia or schizoaffective disorder. *Afr J Psychiatry (Johannesbg).* 11(4), 287–290 (2008).
234. Breen A, Swartz L, Joska J, Flisher AJ, Corrigan J. Adherence to treatment in poorer countries: a new research direction? *Psychiatr Serv.* 58(4), 567–568 (2007).
235. Foussias G, Remington G. Antipsychotics and schizophrenia: from efficacy and effectiveness to clinical decision-making. *Can J Psychiatry.* 55(3), 117–125 (2010).
236. Chisholm D, Saxena S. Cost effectiveness of strategies to combat neuropsychiatric conditions in sub-Saharan Africa and South East Asia: mathematical modelling study. *BMJ.* 344, e609 (2012).
237. Chisholm D, Gureje O, Saldivia S, *et al.* Schizophrenia treatment in the developing world: an interregional and multinational cost-effectiveness analysis. *Bull. World Health Organ.* 86(7), 542–551 (2008).
238. Novick D, Haro JM, Hong J, *et al.* Regional differences in treatment response and three year course of schizophrenia across the world. *J Psychiatr Res.* 46(7), 856–864 (2012).
239. Andrews MM, Boyle JS. *Transcultural Concepts in Nursing Care*, 5e. Lippincott Williams & Wilkins, Philadelphia (2007).

REFERENCES

240. Wright GE, Niehaus DJ, Koen L, Drögemöller BI, Warnich L. Psychiatric genetics in South Africa: cutting a rough diamond. *Afr J Psychiatry (Johannesbg)*. 14(5), 355–366 (2011).
241. Drögemöller BI, Wright GEB, Niehaus DJH, Emsley RA, Warnich L. Whole-genome resequencing in pharmacogenomics: moving away from past disparities to globally representative applications. *Pharmacogenomics*. 12(12), 1717–1728 (2011).
242. Patterson N, Petersen DC, van der Ross RE, *et al*. Genetic structure of a unique admixed population: implications for medical research. *Hum. Mol. Genet.* 19(3), 411–419 (2010).
243. Baye TM, Wilke RA. Mapping genes that predict treatment outcome in admixed populations. *Pharmacogenomics J.* 10(6), 465–477 (2010).
244. Zandi PP, Judy JT. The promise and reality of pharmacogenetics in psychiatry. *Psychiatr. Clin. North Am.* 33(1), 181–224 (2010).
245. Emsley RA, Roberts MC, Rataemane S, *et al*. Ethnicity and treatment response in schizophrenia: a comparison of 3 ethnic groups. *J Clin Psychiatry*. 63(1), 9–14 (2002).
246. Glazer WM, Morgenstern H, Doucette J. Race and tardive dyskinesia among outpatients at a CMHC. *Hosp Community Psychiatry*. 45(1), 38–42 (1994).
247. Holden TJ. Tardive dyskinesia in long-term hospitalised Zulu psychiatric patients. A prevalence study. *S. Afr. Med. J.* 71(2), 88–90 (1987).
248. Chan LF, Zai C, Monda M, *et al*. Role of ethnicity in antipsychotic-induced weight gain and tardive dyskinesia: genes or environment? *Pharmacogenomics*. 14(11), 1273–1281 (2013).
249. Need AC, Goldstein DB. Next generation disparities in human genomics: concerns and remedies. *Trends Genet.* 25(11), 489–494 (2009).
250. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. *Nat. Rev. Genet.* 11(5), 356–366 (2010).
251. Adeyemo A, Gerry N, Chen G, *et al*. A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet.* 5(7), e1000564 (2009).
252. Jallow M, Teo YY, Small KS, *et al*. Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat. Genet.* 41(6), 657–665 (2009).
253. Xiong D-H, Liu X-G, Guo Y-F, *et al*. Genome-wide association and follow-up replication studies identified ADAMTS18 and TGFBR3 as bone mass candidate genes in different ethnic groups. *Am. J. Hum. Genet.* 84(3), 388–398 (2009).
254. Kang SJ, Chiang CWK, Palmer CD, *et al*. Genome-wide association of anthropometric traits in African- and African-derived populations. *Hum. Mol. Genet.* 19(13), 2725–2738 (2010).
255. Thye T, Vannberg FO, Wong SH, *et al*. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat. Genet.* 42(9), 739–741 (2010).
256. Petrovski S, Fellay J, Shianna KV, *et al*. Common human genetic variants and HIV-1 susceptibility: a genome-wide survey in a homogeneous African population. *AIDS*. 25(4), 513–518 (2011).
257. Sirugo G, Hennig BJ, Adeyemo AA, *et al*. Genetic studies of African populations: an overview on disease susceptibility and response to vaccines and therapeutics. *Hum. Genet.* 123(6), 557–598 (2008).
258. Tishkoff SA, Reed FA, Friedlaender FR, *et al*. The genetic structure and history of Africans and African Americans. *Science*. 324(5930), 1035–1044 (2009).
259. Summary by world area | Ethnologue [Internet]. Available from: <http://www.ethnologue.com/statistics/area>.
260. Bryc K, Auton A, Nelson MR, *et al*. Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc. Natl. Acad. Sci. U.S.A.* 107(2), 786–791 (2010).
261. WHO | World Health Organization [Internet]. Available from: http://gamapserver.who.int/gho/interactive_charts/mbd/as_death_rates/atlas.html.

REFERENCES

262. Musunuru K, Strong A, Frank-Kamenetsky M, *et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*. 466(7307), 714–719 (2010).
263. Khoury MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet. Med.* 9(10), 665–674 (2007).
264. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet*. 367(9524), 1747–1757 (2006).
265. Ezeh AC, Izugbara CO, Kabiru CW, *et al.* Building capacity for public and population health research in Africa: the consortium for advanced research training in Africa (CARTA) model. *Glob Health Action*. 3 (2010).
266. Jamison DT. Disease and mortality in sub-Saharan Africa. World Bank Publications, Washington, DC (2006).
267. Kebede S, Duales S, Yokouide A, Alemu W. Trends of major disease outbreaks in the African region, 2003–2007. *East Afr J Public Health*. 7(1), 20–29 (2010).
268. Fiander AN. The prevention of cervical cancer in Africa. *Womens Health (Lond Engl)*. 7(1), 121–132 (2011).
269. Maher D, Sekajugo J. Research on health transition in Africa: time for action. *Health Res Policy Syst*. 9, 5 (2011).
270. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *J Genet Genomics*. 38(3), 95–109 (2011).
271. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology*. 24, 104–108 (1992).
272. Wheeler DA, Srinivasan M, Egholm M, *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 452(7189), 872–876 (2008).
273. Lander ES. QnAs with Eric S. Lander. Interview by Prashant Nair. *Proc. Natl. Acad. Sci. U.S.A.* 108(28), 11319 (2011).
274. Human genome: Genomes by the thousand. *Nature*. 467(7319), 1026–1027 (2010).
275. Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc. Natl. Acad. Sci. U.S.A.* 102(51), 18508–18513 (2005).
276. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet*. 9, 403–433 (2008).
277. About | 1000 Genomes [Internet]. Available from: <http://www.1000genomes.org/about#ProjectSamples>.
278. Bentley DR, Balasubramanian S, Swerdlow HP, *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 456(7218), 53–59 (2008).
279. Kidd JM, Cooper GM, Donahue WF, *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature*. 453(7191), 56–64 (2008).
280. Schuster SC, Miller W, Ratan A, *et al.* Complete Khoisan and Bantu genomes from southern Africa. *Nature*. 463(7283), 943–947 (2010).
281. Abecasis GR, Altshuler D, Auton A, *et al.* A map of human genome variation from population-scale sequencing. *Nature*. 467(7319), 1061–1073 (2010).
282. Caplan A. Genome-sequencing anniversary. What will drive genomics over the next 10 years? *Science*. 331(6018), 691 (2011).
283. Ginsburg GS, Konstance RP, Allsbrook JS, Schulman KA. Implications of pharmacogenomics for drug development and clinical practice. *Arch. Intern. Med.* 165(20), 2331–2336 (2005).
284. Swen JJ, Nijenhuis M, de Boer A, *et al.* Pharmacogenetics: from bench to byte--an update of guidelines. *Clin. Pharmacol. Ther.* 89(5), 662–673 (2011).

REFERENCES

285. Swen JJ, Wilting I, de Goede AL, *et al.* Pharmacogenetics: from bench to byte. *Clin. Pharmacol. Ther.* 83(5), 781–787 (2008).
286. Costa-Scharplatz M, van Asselt ADI, Bachmann LM, Kessels AGH, Severens JL. Cost-effectiveness of pharmacogenetic testing to predict treatment response to angiotensin-converting enzyme inhibitor. *Pharmacogenet. Genomics.* 17(5), 359–368 (2007).
287. Schackman BR, Scott CA, Walensky RP, Losina E, Freedberg KA, Sax PE. The cost-effectiveness of HLA-B*5701 genetic screening to guide initial antiretroviral therapy for HIV. *AIDS.* 22(15), 2025–2033 (2008).
288. Eckman MH, Rosand J, Greenberg SM, Gage BF. Cost-effectiveness of using pharmacogenetic information in warfarin dosing for patients with nonvalvular atrial fibrillation. *Ann. Intern. Med.* 150(2), 73–83 (2009).
289. Gold HT, Hall MJ, Blinder V, Schackman BR. Cost effectiveness of pharmacogenetic testing for uridine diphosphate glucuronosyltransferase 1A1 before irinotecan administration for metastatic colorectal cancer. *Cancer.* 115(17), 3858–3867 (2009).
290. Leey JA, McCabe S, Koch JA, Miles TP. Cost-effectiveness of genotype-guided warfarin therapy for anticoagulation in elderly patients with atrial fibrillation. *Am J Geriatr Pharmacother.* 7(4), 197–203 (2009).
291. Hagaman JT, Kinder BW, Eckman MH. Thiopurine S- methyltransferase [corrected] testing in idiopathic pulmonary fibrosis: a pharmacogenetic cost-effectiveness analysis. *Lung.* 188(2), 125–132 (2010).
292. Vegter S, Jansen E, Postma MJ, Boersma C. Economic evaluations of pharmacogenetic and genomic screening programs: update of the literature. *Drug Development Research.* 71(8), 492–501 (2010).
293. Harnessing Genomic Technologies Toward Improving Health in Africa: Opportunities and Challenges - h3africa_whitepaper.pdf [Internet]. Available from: http://h3africa.org/h3africa_whitepaper.pdf.
294. Mehta U, Durrheim DN, Blockman M, Kredt T, Gounden R, Barnes KI. Adverse drug reactions in adult medical inpatients in a South African hospital serving a community with a high HIV/AIDS prevalence: prospective observational study. *Br J Clin Pharmacol.* 65(3), 396–406 (2008).
295. Wester K, Jönsson AK, Spigset O, Druid H, Hägg S. Incidence of fatal adverse drug reactions: a population based study. *Br J Clin Pharmacol.* 65(4), 573–579 (2008).
296. Microsoft Word - 28 Antiretraviral _ARV_ treatment fact sheet 02 ARV treatment.doc - AntiretraviralARVtreatmentfactsheetARVtreatment.pdf [Internet]. Available from: http://www.haccambodia.org/store_files/book_library/AntiretraviralARVtreatmentfactsheetARVtreatment.pdf.
297. Jones AE, Brown KC, Werner RE, *et al.* Variability in drug metabolizing enzyme activity in HIV-infected patients. *Eur. J. Clin. Pharmacol.* 66(5), 475–485 (2010).
298. Chirac P, Torreele E. Global framework on essential health R&D. *Lancet.* 367(9522), 1560–1561 (2006).
299. Matimba A, Del-Favero J, Van Broeckhoven C, Masimirembwa C. Novel variants of major drug-metabolising enzyme genes in diverse African populations and their predicted functional effects. *Hum. Genomics.* 3(2), 169–190 (2009).
300. Dandara C, Masimirembwa CM, Magimba A, *et al.* Arylamine N-acetyltransferase (NAT2) genotypes in Africans: the identification of a new allele with nucleotide changes 481C>T and 590G>A. *Pharmacogenetics.* 13(1), 55–58 (2003).
301. Wang J, Sönnnerborg A, Rane A, *et al.* Identification of a novel specific CYP2B6 allele in Africans causing impaired metabolism of the HIV drug efavirenz. *Pharmacogenet. Genomics.* 16(3), 191–198 (2006).
302. Gaedigk A, Coetsee C. The CYP2D6 gene locus in South African Coloureds: unique allele distributions, novel alleles and gene arrangements. *Eur. J. Clin. Pharmacol.* 64(5), 465–475 (2008).

REFERENCES

303. Drögemöller BI, Wright GEB, Niehaus DJH, *et al.* Characterization of the genetic profile of CYP2C19 in two South African populations. *Pharmacogenomics*. 11(8), 1095–1103 (2010).
304. Wright GEB, Niehaus DJH, Drögemöller BI, Koen L, Gaedigk A, Warnich L. Elucidation of CYP2D6 genetic diversity in a unique African population: implications for the future application of pharmacogenetics in the Xhosa population. *Ann. Hum. Genet.* 74(4), 340–350 (2010).
305. Mitchell C, Gregersen N, Krause A. Novel CYP2C9 and VKORC1 gene variants associated with warfarin dosage variability in the South African black population. *Pharmacogenomics*. 12(7), 953–963 (2011).
306. Galaxy [Internet]. Available from: <https://main.g2.bx.psu.edu/library>.
307. Thorn CF, Klein TE, Altman RB. Pharmacogenomics and bioinformatics: PharmGKB. *Pharmacogenomics*. 11(4), 501–505 (2010).
308. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31(13), 3812–3814 (2003).
309. Human Cytochrome P450 (CYP) Allele Nomenclature Committee [Internet]. Available from: <http://www.cypalleles.ki.se/>.
310. UGT1A1 allele nomenclature [Internet]. Available from: <http://www.pharmacogenomics.pha.ulaval.ca/files/content/sites/pharmacogenomics/files/Nomenclature/UGT1A/UGT1A1.htm>.
311. Gaedigk A, Simon SD, Pearce RE, Bradford LD, Kennedy MJ, Leeder JS. The CYP2D6 activity score: translating genotype information into a qualitative measure of phenotype. *Clin. Pharmacol. Ther.* 83(2), 234–242 (2008).
312. Ross CJD, Katzov-Eckert H, Dubé M-P, *et al.* Genetic variants in TPMT and COMT are associated with hearing loss in children receiving cisplatin chemotherapy. *Nat. Genet.* 41(12), 1345–1349 (2009).
313. HapMap Homepage [Internet]. Available from: <http://hapmap.ncbi.nlm.nih.gov/>.
314. Rotimi CN, Jorde LB. Ancestry and disease in the age of genomic medicine. *N. Engl. J. Med.* 363(16), 1551–1558 (2010).
315. What to Do with All That Data? | The Daily Scan | GenomeWeb [Internet]. Available from: <http://www.genomeweb.com/blog/what-do-all-data>.
316. Drug Metabolism | Drug Toxicology | Affymetrix [Internet]. Available from: http://www.affymetrix.com/browse/products.jsp?productId=131412&navMode=34000&navAction=jump&ald=productsNav#1_1.
317. Meijerman I, Sanderson LM, Smits PHM, Beijnen JH, Schellens JHM. Pharmacogenetic screening of the gene deletion and duplications of CYP2D6. *Drug Metab. Rev.* 39(1), 45–60 (2007).
318. Gamazon ER, Zhang W, Huang RS, Dolan ME, Cox NJ. A pharmacogene database enhanced by the 1000 Genomes Project. *Pharmacogenet. Genomics*. 19(10), 829–832 (2009).
319. Ng PC, Zhao Q, Levy S, Strausberg RL, Venter JC. Individual genomes instead of race for personalized medicine. *Clin. Pharmacol. Ther.* 84(3), 306–309 (2008).
320. Home - SNP - NCBI [Internet]. Available from: <https://www.ncbi.nlm.nih.gov/snp/>.
321. Hayden E. Human genome at ten: Life is complicated. *Nature*. 464(7289), 664–667 (2010).
322. Frommolt P, Abdallah AT, Altmüller J, *et al.* Assessing the enrichment performance in targeted resequencing experiments. *Hum. Mutat.* 33(4), 635–641 (2012).
323. Kirby A, Gnirke A, Jaffe DB, *et al.* Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. *Nat. Genet.* (2013).
324. Altman RB, Whirl-Carrillo M, Klein TE. Challenges in the pharmacogenomic annotation of whole genomes. *Clin. Pharmacol. Ther.* (2013).
325. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25(14), 1754–1760 (2009).
326. DePristo MA, Banks E, Poplin R, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43(5), 491–498 (2011).

REFERENCES

327. Index of /vol1/ftp/phase1/analysis_results/supporting/accessible_genome_masks [Internet]. Mask files. Available from: http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/supporting/accessible_genome_masks/.
328. Warren RL, Choe G, Freeman DJ, *et al.* Derivation of HLA types from shotgun sequence datasets. *Genome Med.* 4(12), 95 (2012).
329. Babik W, Taberlet P, Ejsmond MJ, Radwan J. New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Mol Ecol Resour.* 9(3), 713–719 (2009).
330. HLA Nomenclature @ hla.alleles.org [Internet]. Available from: <http://hla.alleles.org/>.
331. Ingelman-Sundberg M. The human genome project and novel aspects of cytochrome P450 research. *Toxicol. Appl. Pharmacol.* 207(2 Suppl), 52–56 (2005).
332. Ensembl: BioMart [Internet]. Available from: <http://www.ensembl.org/biomart/martview/e25e6465bdfc8ee9b691003eca91ab67>.
333. Home - Gene - NCBI [Internet]. Available from: <http://www.ncbi.nlm.nih.gov/gene>.
334. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32(Web Server issue), W273–279 (2004).
335. RepeatMasker Home Page [Internet]. Available from: <http://www.repeatmasker.org/>.
336. The R Project for Statistical Computing [Internet]. Available from: <http://www.r-project.org/>.
337. Jafari S, Fernandez-Enright F, Huang X-F. Structural contributions of antipsychotic drugs to their therapeutic profiles and metabolic side effects. *J. Neurochem.* 120(3), 371–384 (2012).
338. Røge R, Møller BK, Andersen CR, Correll CU, Nielsen J. Immunomodulatory effects of clozapine and their clinical implications: what have we learned so far? *Schizophr. Res.* 140(1-3), 204–213 (2012).
339. Kiezun A, Garimella K, Do R, *et al.* Exome sequencing and the genetic basis of complex traits. *Nat. Genet.* 44(6), 623–630 (2012).
340. Duan J, Zhang J-G, Deng H-W, Wang Y-P. Comparative studies of copy number variation detection methods for next-generation sequencing technologies. *PLoS ONE.* 8(3), e59128 (2013).
341. Network-wide Projects - PGRN Web Site - Confluence [Internet]. Available from: <http://pgrn.org/display/pgrnwebsite/Network-wide+Projects>.
342. Gaedigk A, Ndjountché L, Divakaran K, *et al.* Cytochrome P4502D6 (CYP2D6) gene locus heterogeneity: characterization of gene duplication events. *Clin. Pharmacol. Ther.* 81(2), 242–251 (2007).
343. Oxford Nanopore Technologies [Internet]. Available from: <http://www.nanoporetech.com/>.
344. Manolio TA, Collins FS, Cox NJ, *et al.* Finding the missing heritability of complex diseases. *Nature.* 461(7265), 747–753 (2009).
345. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell.* 148(6), 1223–1241 (2012).
346. Genomic Services Laboratory at HudsonAlpha [Internet]. Available from: <http://www.hudsonalpha.org/gsl/index>.
347. Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 25(16), 2078–2079 (2009).
348. SeattleSeq Variation Annotation [Internet]. Available from: <http://snp.gs.washington.edu/SeattleSeqAnnotation134/>.
349. 1000 Genomes [Internet]. Available from: <http://browser.1000genomes.org/index.html>.
350. Putnam DK, Sun J, Zhao Z. Exploring schizophrenia drug-gene interactions through molecular network and pathway modeling. *AMIA Annu Symp Proc.* 2011, 1127–1133 (2011).
351. Lim ET, Raychaudhuri S, Sanders SJ, *et al.* Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron.* 77(2), 235–242 (2013).

REFERENCES

352. Smoller JW, Craddock N, Kendler K, *et al.* Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*. 381(9875), 1371–1379 (2013).
353. Brunham LR, Hayden MR. Hunting human disease genes: lessons from the past, challenges for the future. *Hum Genet*. 132(6), 603–617 (2013).
354. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 16(3), 1215 (1988).
355. Yandell M, Huff C, Hu H, *et al.* A probabilistic disease-gene finder for personal genomes. *Genome Res*. 21(9), 1529–1542 (2011).
356. Adzhubei IA, Schmidt S, Peshkin L, *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods*. 7(4), 248–249 (2010).
357. Daya M, van der Merwe L, Galal U, Möller M, Salie M, Chimusa, Emile. A panel of ancestry informative markers for the complex five-way admixed South African Coloured population. *PLoS One*. In Press.
358. HSC Cores Research Facility – Genomics [Internet]. Available from: http://www.cores.utah.edu/?page_id=3655.
359. Miller M. Tools for population genetic analyses (TFPGA) 1.3: a Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by the author. (1997).
360. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 19(9), 1655–1664 (2009).
361. CRAN - Package genetics [Internet]. Available from: <http://cran.r-project.org/web/packages/genetics/index.html>.
362. Emond MJ, Louie T, Emerson J, *et al.* Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *Nat. Genet*. 44(8), 886–889 (2012).
363. Nho K, Corneveaux JJ, Kim S, *et al.* Whole-exome sequencing and imaging genetics identify functional variants for rate of change in hippocampal volume in mild cognitive impairment. *Mol. Psychiatry*. 18(7), 781–787 (2013).
364. Di Fusco D, Ciccacci C, Rufini S, Forte V, Novelli G, Borgiani P. Resequencing of VKORC1, CYP2C9 and CYP4F2 genes in Italian patients requiring extreme low and high warfarin doses. *Thromb. Res*. 132(1), 123–126 (2013).
365. Statistics South Africa - Home [Internet]. Available from: <http://www.statssa.gov.za/publications/populationstats.asp>.
366. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet*. (2013).
367. Frosst P, Blom HJ, Milos R, *et al.* A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat. Genet*. 10(1), 111–113 (1995).
368. Lajin B, Alhaj Sakur A, Michati R, Alachkar A. Association between MTHFR C677T and A1298C, and MTRR A66G polymorphisms and susceptibility to schizophrenia in a Syrian study cohort. *Asian J Psychiatr*. 5(2), 144–149 (2012).
369. Allen NC, Bagade S, McQueen MB, *et al.* Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat. Genet*. 40(7), 827–834 (2008).
370. Roffman JL, Brohawn DG, Nitenson AZ, Macklin EA, Smoller JW, Goff DC. Genetic variation throughout the folate metabolic pathway influences negative symptom severity in schizophrenia. *Schizophr Bull*. 39(2), 330–338 (2013).
371. MTRR - PubMed - NCBI [Internet]. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/?term=MTRR>.
372. Joober R, Benkelfat C, Lal S, *et al.* Association between the methylenetetrahydrofolate reductase 677C-->T missense mutation and schizophrenia. *Mol. Psychiatry*. 5(3), 323–326 (2000).

REFERENCES

373. Vehof J, Burger H, Wilffert B, *et al.* Clinical response to antipsychotic drug treatment: association study of polymorphisms in six candidate genes. *Eur Neuropsychopharmacol.* 22(9), 625–631 (2012).
374. Niehaus DJH, Koen L, Laurent C, *et al.* Positive and negative symptoms in affected sib pairs with schizophrenia: implications for genetic studies in an African Xhosa sample. *Schizophr. Res.* 79(2-3), 239–249 (2005).
375. Wright GEB, Niehaus DJH, van der Merwe L, *et al.* Association of MB-COMT polymorphisms with schizophrenia-susceptibility and symptom severity in an African cohort. *Prog. Neuropsychopharmacol. Biol. Psychiatry.* 39(1), 163–169 (2012).
376. SPAG17 sperm associated antigen 17 [Homo sapiens (human)] - Gene - NCBI [Internet]. Available from: <http://www.ncbi.nlm.nih.gov/gene/?term=SPAG17>.
377. DNHD1 dynein heavy chain domain 1 [Homo sapiens (human)] - Gene - NCBI [Internet]. Available from: <http://www.ncbi.nlm.nih.gov/gene/144132>.
378. Tervaniemi MH, Siitonen HA, Söderhäll C, *et al.* Centrosomal localization of the psoriasis candidate gene product, CCHCR1, supports a role in cytoskeletal organization. *PLoS ONE.* 7(11), e49920 (2012).
379. Kott E, Duquesnoy P, Copin B, *et al.* Loss-of-function mutations in LRRC6, a gene essential for proper axonemal assembly of inner and outer dynein arms, cause primary ciliary dyskinesia. *Am. J. Hum. Genet.* 91(5), 958–964 (2012).
380. Mitchison HM, Schmidts M, Loges NT, *et al.* Mutations in axonemal dynein assembly factor DNAAF3 cause primary ciliary dyskinesia. *Nat. Genet.* 44(4), 381–389, S1–2 (2012).
381. DNAAF1 dynein, axonemal, assembly factor 1 [Homo sapiens (human)] - Gene - NCBI [Internet]. Available from: <https://www.ncbi.nlm.nih.gov/gene/123872>.
382. Sakakibara A, Ando R, Sapir T, Tanaka T. Microtubule dynamics in neuronal morphogenesis. *Open Biol.* 3(7), 130061 (2013).
383. Catts VS, Fung SJ, Long LE, *et al.* Rethinking schizophrenia in the context of normal neurodevelopment. *Front Cell Neurosci.* 7, 60 (2013).
384. Peterson BS. Neuroimaging in child and adolescent neuropsychiatric disorders. *J Am Acad Child Adolesc Psychiatry.* 34(12), 1560–1576 (1995).
385. Bunney BG, Potkin SG, Bunney WE Jr. New morphological and neuropathological findings in schizophrenia: a neurodevelopmental perspective. *Clin. Neurosci.* 3(2), 81–88 (1995).
386. Fatemi SH, Folsom TD. The neurodevelopmental hypothesis of schizophrenia, revisited. *Schizophr Bull.* 35(3), 528–548 (2009).
387. Jaglin XH, Chelly J. Tubulin-related cortical dysgeneses: microtubule dysfunction underlying neuronal migration defects. *Trends Genet.* 25(12), 555–566 (2009).
388. Reiner O. LIS1 and DCX: Implications for brain development and human disease in relation to microtubules. *Scientifica.* 2013, 393975 (2013).
389. Asumalahti K, Laitinen T, Itkonen-Vatjus R, *et al.* A candidate gene for psoriasis near HLA-C, HCR (Pg8), is highly polymorphic with a disease-associated susceptibility allele. *Hum. Mol. Genet.* 9(10), 1533–1542 (2000).
390. Moskvina V, Smith M, Ivanov D, *et al.* Genetic differences between five European populations. *Hum. Hered.* 70(2), 141–149 (2010).
391. False Positives from Next-Gen Sequencing | MassGenomics [Internet]. Available from: <http://massgenomics.org/2013/06/ngs-false-positives.html>.
392. EST Profile - Hs.415342 [Internet]. Available from: <http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi?uglist=Hs.415342>.
393. Insel TR. Rethinking schizophrenia. *Nature.* 468(7321), 187–193 (2010).
394. H3Africa | Homepage [Internet]. Available from: <http://h3africa.org/>.
395. Ensembl Genome Browser [Internet]. Available from: <http://www.ensembl.org/index.html>.
396. mVISTA Input [Internet]. Available from: <http://genome.lbl.gov/vista/mvista/submit.shtml>.

REFERENCES

397. De Wit E, Delpont W, Rugamika CE, *et al.* Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. *Hum. Genet.* 128(2), 145–153 (2010).
398. Suarez-Kurtz G. Pharmacogenetics in the Brazilian population. *Front Pharmacol.* 1, 118 (2010).
399. Hinds DA, Stuve LL, Nilsen GB, *et al.* Whole-genome patterns of common DNA variation in three human populations. *Science.* 307(5712), 1072–1079 (2005).
400. De Candia TR, Lee SH, Yang J, *et al.* Additive Genetic Variation in Schizophrenia Risk Is Shared by Populations of African and European Descent. *Am. J. Hum. Genet.* (2013).
401. Liu L, Li Y, Li S, *et al.* Comparison of next-generation sequencing systems. *BioMed Research International.* 2012, 251364 (2012).
402. Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. *Genome Biology.* 14(6), 405 (2013).
403. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res.* 22(9), 1748–1759 (2012).
404. Wright GEB, Koornhof PGJ, Adeyemo AA, Tiffin N. Ethical and legal implications of whole genome and whole exome sequencing in African populations. *BMC Med Ethics.* 14, 21 (2013).
405. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med.* 11, 126 (2013).
406. Genetic Analysis | SNP Analysis | Affymetrix [Internet]. Available from: http://www.affymetrix.com/estore/browse/products.jsp?productId=131533#1_1.
407. Boyadjieva N, Varadinova M. Epigenetics of psychoactive drugs. *J. Pharm. Pharmacol.* 64(10), 1349–1358 (2012).
408. Avramopoulos D. Genetics of psychiatric disorders methods: molecular approaches. *Clin. Lab. Med.* 30(4), 815–827 (2010).
409. Lin M, Pedrosa E, Shah A, *et al.* RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS ONE.* 6(9), e23356 (2011).
410. Brennand KJ, Gage FH. Concise review: the promise of human induced pluripotent stem cell-based studies of schizophrenia. *Stem Cells.* 29(12), 1915–1922 (2011).
411. Im H-I, Kenny PJ. MicroRNAs in neuronal function and dysfunction. *Trends Neurosci.* 35(5), 325–334 (2012).
412. Kaffman A, Krystal JH, Krystal JJ. New frontiers in animal research of psychiatric illness. *Methods Mol. Biol.* 829, 3–30 (2012).
413. Kannan G, Sawa A, Pletnikov MV. Mouse models of gene-environment interactions in schizophrenia. *Neurobiol. Dis.* 57C, 5–11 (2013).
414. Brennand KJ, Simone A, Jou J, *et al.* Modelling schizophrenia using human induced pluripotent stem cells. *Nature.* 473(7346), 221–225 (2011).
415. Rado J, Janicak PG. Iloperidone for schizophrenia. *Expert Opin Pharmacother.* 11(12), 2087–2093 (2010).
416. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
417. FileZilla - The free FTP solution [Internet]. Available from: <https://filezilla-project.org/>.
418. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams [Internet]. Available from: <http://www.cmbi.ru.nl/cdd/biovenn/index.php>.
419. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 21(2), 263–265 (2005).
420. IDT - PrimerQuest Input [Internet]. Available from: <http://eu.idtdna.com/PrimerQuest/Home/Index>.
421. Oligo Analyzer [Internet]. Available from: <http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/>.

REFERENCES

422. Primer designing tool [Internet]. Available from: http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome.
423. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series*. 41, 95–98 (1999).
424. RestrictionMapper version 3 [Internet]. Available from: <http://www.restrictionmapper.org/>.
425. Genomics [Internet]. Available from: <http://www.cores.utah.edu/maincores/main-core-facilities/genomics>.