Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Pharmacogenomics

Fine-mapping of antipsychotic response genome-wide association studies reveals novel regulatory mechanisms

Aim: Noncoding variation has demonstrated regulatory effects on disease treatment outcomes. This study investigated the potential functionality of previously implicated noncoding variants on schizophrenia treatment response. **Materials & methods:** Predicted regulatory potential of variation identified from antipsychotic response genome-wide association studies was determined. Prioritized variants were assessed for association(s) with treatment outcomes in a South African first episode schizophrenia cohort (n = 103). **Results:** Bioinformatic and association results implicated a relationship between regulatory variants, expression of *MANBA*, *COL9A2* and *NFKB1*, and treatment response. Three SNPs were associated with poor outcomes (rs230493: $p = 1.88 \times 10^{-6}$; rs3774959: $p = 1.75 \times 10^{-5}$; and rs230504: $p = 1.48 \times 10^{-4}$). **Conclusion:** This study has thoroughly investigated previous GWAS to pinpoint variants that may play a causal role in poor schizophrenia treatment outcomes, and provides potential candidate genes for further study in the field of antipsychotic response.

Ellen S Ovenden[1], Britt I Drögemöller[†,1], Lize van der Merwe[1], Bonginkosi Chiliza[2], Laila Asmal[2], Robin A Emsley[2] & Louise Warnich*[,1]

[1]Department of Genetics, Stellenbosch University, Stellenbosch, South Africa

[2]Department of Psychiatry, Stellenbosch University, Tygerberg, South Africa

*Author for correspondence: lw@sun.ac.za

[†]Current address: Department of Pediatrics, University of British Columbia, British Columbia, Canada

First draft submitted: 19 June 2016; Accepted for publication: 30 September 2016; Published online: 19 December 2016

Keywords: antipsychotic treatment response • *COL9A2* • ENCODE • flupenthixol decanoate • *MANBA* • *NFKB1* • rSNPs • schizophrenia

Schizophrenia is a severe psychiatric disorder, usually requiring long-term maintenance antipsychotic treatment. Treatment response varies widely, with many patients experiencing persistent symptoms despite ongoing treatment. The extensive heterogeneity observed in treatment response is largely brought about by the genetic differences between individuals in drug metabolism, neurotransmission and other pathways [1–3]. However, studies to date have mostly provided inconsistent results with regard to candidate genes. Much like other complex disorders, there are likely to be hundreds to thousands of common variants across the genome that cumulatively contribute to individual treatment response phenotypes, including adverse drug reactions (ADRs) [4]. With large and/or well-characterized sample groups, pharmacogenomics enables the discovery of these variants. Specifically, genome-wide association studies (GWAS) have recently been employed to study antipsychotic response. While this approach improves upon *a priori* candidate gene studies by analyzing variants across the genome, the majority of GWAS lack sufficient biological interpretation. Often, variants are considered in isolation and exclusively with regard to the function of their neighboring gene [5,6]. This restricts the creation of new hypotheses and further understanding of treatment response mechanisms.

Traditionally, coding SNPs have proved far more amenable to functional analyses, which means that they are often prioritized post-GWAS for further study, while noncoding variants are ignored [7]. Recently,

Future Medicine part of fsg

we have learnt a great deal more about noncoding regions, largely due to the plethora of results from the ENCODE project, which sought to characterize all the functional elements of the genome, including regulatory factors [8]. With this knowledge, it has been revealed that noncoding SNPs implicated in regulation, or rSNPs, are enriched for in GWAS associations, highlighting the importance of analyzing these regions for implications in disease and drug response [9,10]. Many bioinformatic tools that make use of the abundance of ENCODE data have recently been developed. For example, RegulomeDB [11] and rSNPBase [12] both assess the regulatory potential of a SNP: RegulomeDB looks at expression quantitative trait loci (eQTL) evidence and proximal regulation, and rSNPBase assesses proximal, distal and post-transcriptional functioning and predicts downstream gene targets.

Considering these factors, this study aimed to investigate the functionality of noncoding variants linked to antipsychotic response, through bioinformatic and association analyses. We assessed previous antipsychotic response GWAS coupled with linkage disequilibrium (LD) analyses, and investigated functionality with the use of recent *in silico* tools. A novel bioinformatics approach was developed, in which experimentally validated ENCODE data as well as predictive tools were employed to isolate rSNPs and their regulatory impact. To validate these results, the top variants were genotyped in a well-characterized South African cohort of first episode schizophrenia (FES) patients, and assessed for associations with treatment outcomes.

Materials & methods

Bioinformatic analyses

Data mining

The literature was mined to identify all variants from GWAS that have been significantly associated with antipsychotic response in schizophrenia, including ADRs. This was accomplished with the NHGRI GWAS Catalog [13,14], last accessed in August 2014. An acceptable but less stringent genome-wide significance level of $p \leq 5 \times 10^{-7}$ [15] was employed to filter SNPs, due to the limited amount of studies and in the interest of being as inclusive as possible. Furthermore, the majority of psychiatric pharmacogenetic studies lack the sample size needed to reach the commonly used 5×10^{-8} threshold.

Regions of linkage disequilibrium

Variants in high LD ($r^2 \geq 0.8$) [16] with the associated GWAS variants were identified using SNP Annotation and Proxy search (SNAP) version 2.2 [17]. Different population groups were analyzed depending on the ancestral make-up of the patient samples in the GWAS. The analyses included four data sets (latest versions available at the time of analysis): 1000 Genomes Pilot 1, HapMap Phase II release 21, HapMap Phase II release 22 and HapMap Phase III release 2 [18,19]. For each GWAS variant, only SNPs in LD in all relevant population groups and across both 1000 Genomes and HapMap datasets were included for further analyses.

rSNP prediction

To assess the potential regulatory impact of the implicated regions, the list of variants was uploaded to RegulomeDB (Version 1) [20]. With the use of a heuristic scoring system, RegulomeDB predicts the degree to which an intergenic SNP will interfere with binding and downstream regulatory processes (see [11] for an explanation of the scoring system). SNPs with a score of ≤ 3 were considered important for further investigation [21]. Additionally, rSNPBase [22] was also employed to assess the regulatory potential of the variants and to determine the predicted affected gene targets. Targets are obtained from experimentally supported databases that link a particular SNP to a gene via a regulatory process.

Variant prioritization

Regulatory SNPs were prioritized for genotyping in our cohort according to the RegulomeDB results (score ≤3). These findings are robust since they are based on experimentally validated data such as ENCODE [11]. rSNPBase does not use a scoring system to rank variants, however extensive overlap between the top RegulomeDB variants and rSNPs predicted by rSNPBase was observed. RegulomeDB variants were coupled with their proxy SNPs from antipsychotic response GWAS (provided the GWAS SNP was not already classified as an rSNP) for a set of 30 variants (Supplementary Table 1). Figure 1 illustrates an overview of the prioritization process .

Genetic association analyses

Patient samples

A South African FES cohort of 103 patients (median age: 23 ± 7 years; 74% male) was used to investigate associations with SNPs previously implicated in treatment response [23]. The cohort consisted of 82 South African Colored (SAC), 13 Xhosa and eight Caucasian individuals. Patients were recruited over four years at Stikland Hospital in the Western Cape and assessed with the Structured Clinical Interview for the DSM-IV [24]. Written informed consent was provided by all patients or their caregivers prior to the study. Ethical approval was obtained from the Human Research and Ethics Committee (HREC), Faculty of

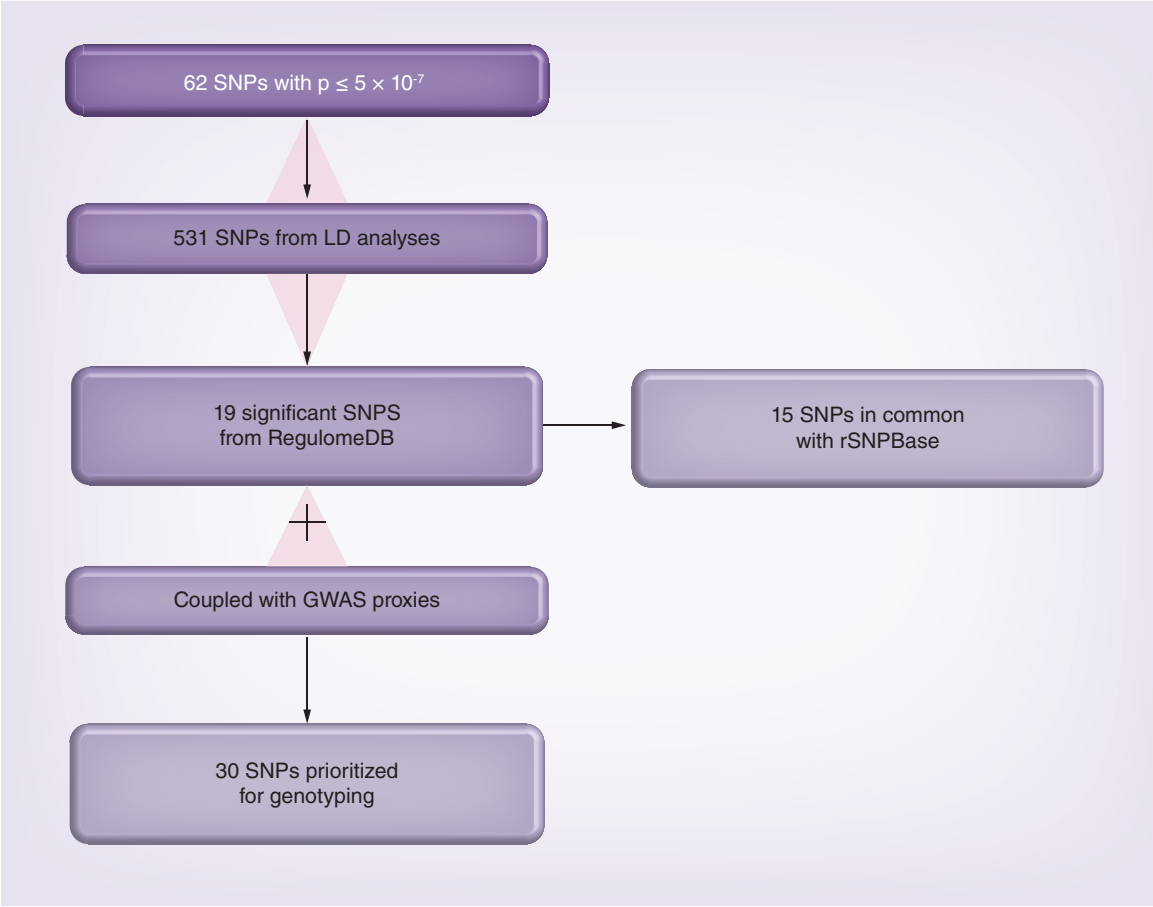


Figure 1. Schematic illustrating the number of significant variants after each step of the bioinformatics approach, resulting in variant prioritization for genotyping.
GWAS: Genome-wide association studies; LD: Linkages disequilibrium.

Medicine and Health Sciences, Stellenbosch University (ethics numbers for clinical and genetic aspects: N06/08/148 and 1907/005, respectively).

All patients received treatment with flupenthixol decanoate (Fluanxol®, Lundbeck, Copenhagen, Denmark) by long-acting injection, according to a fixed protocol [23]. Dose was initiated at 10 mg 2-weekly and gradually increased until remission was achieved – defined according to the Remission in Schizophrenia Working Group [25] – or until the maximum recommended or tolerated dose was reached. Response to treatment was measured by the Positive and Negative Syndrome Scale (PANSS) [26] over a period of 12 months, with assessments conducted every 2 weeks for the first 6 weeks, and every 3 months thereafter. Treatment outcomes assessed in this study were percentage changes per month in PANSS scores in the four PANSS domains, in other words, positive, negative, general and total symptoms.

Prior to the study, genomic DNA (gDNA) was extracted from whole blood samples from each patient as previously described [27,28].

Variant genotyping

TaqMan® OpenArray® Real-Time PCR (Life Technologies™, NY, USA) was employed to genotype the 30 SNPs (Supplementary Table 1) in the cohort. TaqMan assays were obtained from the SNP Genotyping Assay Search Tool [29]. In cases where no predesigned assay was available, a custom assay was designed, and the final 30-SNP assay was manufactured by Life Technologies. SNPs that failed custom assay design or functional testing were excluded and replaced by the next most significant RegulomeDB SNP and GWAS proxy. Genotyping of DNA samples was performed at the University of Utah DNA Sequencing and Genomics Core Facility (UT, USA). Duplicate samples were included to assess reproducibility of the assay, and two wells without template DNA per plate served as negative controls.

Statistical analyses

Allele and genotype frequencies for successfully genotyped SNPs were calculated, and Hardy–Weinberg equilibrium (HWE) was tested by means of a Pearson's

Chi-square (χ^2) test or analogue to Fisher's Exact Test with SNPStats [30,31]. SNPs with $p < 0.01$ were considered to deviate from HWE. Subsequently, LD between SNPs ($r^2 \geq 0.8$; LOD ≥3) was assessed with Haploview version 4.2 [32].

All statistical modeling was done with functions from base R and from various R packages. Pairs of haplotypes, together with their probabilities of being correct, were inferred for each individual using functions from the R package *haplo.stats* [33]. Linear mixed-effects modeling was done using the R packages *nlme* [34], *lme4* [35] and *lmerTest* [36]. The four types of log-transformed (to attain approximate normality) PANSS scores (positive, negative, general and total) over time were modeled as functions of the interaction between time of observation and genetic factor as fixed effects. A random effect was included to correct for the correlations between multiple measurements over time on the same individual. The genetic predictors in the models were genotypes (two degrees of freedom), additive allelic (counting number of minor alleles) and counting individual number of haplotypes, respectively. When a significant allelic or genotype effect was detected, we inspected the estimates and then tested the association with the most likely of dominant, recessive or heterozygote models. Only the results from the specific model that fitted the observed data best (smallest p-value) are reported.

Since the SAC population is highly admixed, any spurious associations due to population stratification were accounted for by adjusting each model for ancestry contributions of five different ancestral groups. This was accomplished by utilizing previously genotyped ancestry informative markers as covariates to estimate ancestry proportions in ADMIXTURE [37–39]. The ancestry proportions of the 82 SAC individuals are indicated in Supplementary Figure 1. In addition to proportion ancestry, all models were also adjusted for baseline PANSS scores, age, sex and self-reported ethnic group (Caucasian, SAC or Xhosa). Effect sizes and their 95% CIs were derived from these models. Because the PANSS scores were log-transformed, effect sizes are estimated percentage changes over time.

Bonferroni was used to correct for multiple testing after association analyses. The significance threshold of 0.05 was divided by the total number of tests (n = 240) performed in the initial analyses. The number of tests was calculated as follows: number of tests = (number of variants [n = 30]) × (number of models used [n = 2, genotypic and additive allelic]) × (number of phenotypes [n = 4, PANSS positive, negative, total and general scores]). Therefore the adjusted significance threshold is 0.0002 (0.05/240).

Results

Bioinformatic analyses

Figure 1 illustrates the process of SNP selection and prioritization using the bioinformatics approach.

Data mining & LD analyses

Data mining of literature uncovered nine GWAS that identified a combined total of 62 SNPs associated with antipsychotic treatment responses in schizophrenia at $p \leq 5 \times 10^{-7}$ (Table 1). The associated responses were diverse and included: adverse motor side effects, metabolic changes, changes in symptom severity rated both by the patient and clinician, adverse cardiac symptoms and treatment-refractoriness. Six of the nine GWAS made use of the data from the CATIE trial [40] to identify associations.

For the LD analyses, populations were selected according to the ancestral make-up of each GWAS, as indicated in Supplementary Table 2. All GWAS were corrected for ancestry, in cases where the cohort was comprised of more than one ethnic group. Additionally, many of the studies performed subsample testing in order to ascertain whether a particular population group was driving a significant association. Except for a few instances in which an SNP was invariant in a particular subgroup, the majority of SNPs were significant – albeit to varying degrees – across population subgroups. Therefore, to be as inclusive as possible, no population groups were excluded for further LD analysis. Once results from different population groups were combined, there was a total of 531 unique SNPs.

rSNP prediction

Nineteen SNPs received scores meeting the RegulomeDB threshold set in this study (≤3). These include two original GWAS SNPs, namely rs6741819 [42] and rs10458561 [5], and 17 SNPs identified from LD analyses. All 19 SNPs show evidence for changes in chromatin state and histone modifications at their particular locus.

The top nine results obtained a RegulomeDB score of 1, meaning that they have shown to act as eQTLs, altering the expression of several genes. These variants are listed in Table 2 together with their predicted regulatory effect. All nine SNPs are in LD with a significant variant (rs230529) from the study by Liou *et al.* [46] and are all eQTLs for both *MANBA* gene and the *COL9A2* gene. These nine variants also form part of a group of 14 SNPs that occur in the 4q24 region. Further investigation of this locus with the use of the University of California, Santa Cruz (UCSC) ENCODE browser [48] revealed that there are several

Table 1. Significant SNPs from antipsychotic pharmacogenomic GWAS identified by the NHGRI GWAS Catalog.

Study (year)	Sample		Variant	Locus	Gene	Response measurement	p-value	Effect†	Ref.
	Size: initial; replication	Ethnicity							
Åberg *et al.* (2010)	738‡	57% EA, 29% AA, 14% other	rs17022444	2p12	Intergenic	EPS (SAS)	1×10^{-10}	+	[41]
			rs7669317	4q24	Intergenic	EPS (AIMS)	8×10^{-8}	+	
			rs2126709	11q24.1	*ZNF202* (3′-UTR)	EPS (SAS)	4×10^{-7}	+	
Adkins *et al.* (2011)	738‡	57% EA, 29% AA, 14% other	rs1568679	15q14	*MEIS2* (intron)	Hip circumference	1×10^{-8}	+	[42]
			rs1967256	5q14.3	*GPR98* (intron)	Hemoglobin A1c	3×10^{-8}	+	
			rs11954387	5q14.3	*GPR98* (intron)	Hemoglobin A1c	3×10^{-8}	+	
			rs1405687	4q24	Intergenic	Hip circumference	5×10^{-8}	-	
			rs1568679	15q14	*MEIS2* (intron)	Waist circumference	6×10^{-8}	+	
			rs13224682	7p22.3	*PRKAR2B* (intron)	Triglycerides	6×10^{-8}	+	
			rs1464500	12p12.1	*SOX5* (intron)	HDL cholesterol	1×10^{-7}	+	
			rs17651157	18q12.2	*FHOD3* (intron)	Triglycerides	1×10^{-7}	+	
			rs6735179	2p25.3	Intergenic	Triglycerides	1×10^{-7}	+	
			rs518590	13q12.11	Intergenic	HDL cholesterol	2×10^{-7}	+	
			rs10502661	18q12.2	*FHOD3* (intron)	Triglycerides	2×10^{-7}	+	
			rs1187614	14q32.13	*CLMN* (intron)	Total cholesterol	2×10^{-7}	-	
			rs6741819	2p25.1	*RNF144A* (intron)	Triglycerides	2×10^{-7}	+	
			rs4838255	9q33.1	*ASTN2* (intron)	Triglycerides	3×10^{-7}	+	
			rs2994684	10p11.22	Intergenic	Triglycerides	3×10^{-7}	+	
			rs977396	8q22.3	Intergenic	Total cholesterol	3×10^{-7}	+	
			rs7105881	11q23.1	Intergenic	Hip circumference	3×10^{-7}	+	
			rs1117324	2p24.1	Intergenic	Hip circumference	3×10^{-7}	+	
			rs4783227	16q23.3	Intergenic	Total cholesterol	4×10^{-7}	-	
			rs320209	9q31.1	Intergenic	Glucose	4×10^{-7}	+	
			rs7108821	11q23.1	Intergenic	Hip circumference	4×10^{-7}	+	
			rs10499504	7p21.1	Intergenic	Total cholesterol	4×10^{-7}	-	
			rs7119817	11q23.1	Intergenic	Hip circumference	5×10^{-7}	+	

†Direction of effect of minor allele, where '+' denotes minor allele frequency associated with poorer response or presence of ADRs.
‡Identical cohort from the CATIE study [40].
§Joint probability from meta-analysis of initial and replication cohorts.
AA: African–American; ADR: Adverse drug reaction; AIMS: Abnormal involuntary movement scale; Cau: Caucasian; CGI-S: Clinical global impression severity scale; EA: European–American; EPS: Extrapyramidal symptom; Hemoglobin A1c: Glycohemoglobin (used to measure plasma glucose levels); HDL: High-density lipoprotein; PANSS: Positive and negative syndrome scale; PGI: Patient global impression scale; QTc: Interval between ventricular depolarization (Q wave) and repolarization (T wave) in electrocardiogram, corrected for heart rate; SAS: Simpson–Angus scale.

Table 1. Significant SNPs from antipsychotic pharmacogenomic GWAS identified by the NHGRI GWAS Catalog (cont.).

Study (year)	Sample		Variant	Locus	Gene	Response measurement	p-value	Effect†	Ref.
	Size: initial; replication	Ethnicity							
Adkins *et al.* (2011; cont.)			rs9658108	6p21.31	*PPARD* (intron)	Glucose	5×10^{-7}	+	
			rs17100498	5q31.3	Intergenic	Hemoglobin A1c	5×10^{-7}	+	
			rs399885	2p12	Intergenic	Heart rate	5×10^{-7}	+	
McClay *et al.* (2011)	738‡	57% EA, 29% AA, 14% other	rs286913	11p13	*EHF* (intron)	Neurocognition: vigilance	7×10^{-8}	-	[43]
			rs11240594	1q32.1	*SLC26A9* (intron)	Neurocognition: processing speed	1×10^{-7}	-	
			rs11110077	12q23.1	*ANKS1B* (intron)	Neurocognition: working memory	4×10^{-7}	+	
			rs7520258	1q42.3	*GPR137B* (intron)	Neurocognition: working memory	5×10^{-7}	+	
			rs12726652	1p13.3	Intergenic	Neurocognition: working memory	5×10^{-7}	+	
			rs11214606	11q23.2	*DRD2* (intron)	Neurocognition: working memory	5×10^{-7}	+	
			rs2833556	21q22.11	*HUNK* (intron)	Neurocognition: reasoning	5×10^{-7}	-	
McClay *et al.* (2011)	738‡	57% EA, 29% AA, 14% other	rs17390445	4p15.1	Intergenic	Positive symptoms (PANSS)	1×10^{-7}	+	[44]
			rs888219	9q33.3	Intergenic	Negative symptoms (PANSS)	2×10^{-7}	-	
			rs7968606	12q23.1	*ANKS1B* (intron)	Negative symptoms (PANSS)	3×10^{-7}	-	
			rs17727261	2q14.3	*CNTNAP5* (exon)	Negative symptoms (PANSS)	5×10^{-7}	-	
			rs11722719	4p15.1	Intergenic	Positive symptoms (PANSS)	5×10^{-7}	+	
Åberg *et al.* (2012)	738‡	57% EA, 29% AA, 14% other	rs4959235	6p25.2	*SLC22A23* (intron)	QTc interval prolongation	2×10^{-7}	+	[5]
			rs10458561	1p31.1	Intergenic		4×10^{-7}	+	
Athanasiu *et al.* (2012)	594	100% Cau	rs7838490	8q21.3	Intergenic	BMI	6×10^{-8}	+	[45]
			rs11615274	12q21.1	Intergenic	HDL cholesterol	9×10^{-8}	-	

†Direction of effect of minor allele, where '+' denotes minor allele frequency associated with poorer response or presence of ADRs.
‡Identical cohort from the CATIE study [40].
§Joint probability from meta-analysis of initial and replication cohorts.
AA: African–American; ADR: Adverse drug reaction; AIMS: Abnormal involuntary movement scale; Cau: Caucasian; CGI-S: Clinical global impression severity scale; EA: European–American; EPS: Extrapyramidal symptom; Hemoglobin A1c: Glycohemoglobin (used to measure plasma glucose levels); HDL: High-density lipoprotein; PANSS: Positive and negative syndrome scale; PGI: Patient global impression scale; QTc: Interval between ventricular depolarization (Q wave) and repolarization (T wave) in electrocardiogram, corrected for heart rate; SAS: Simpson–Angus scale.

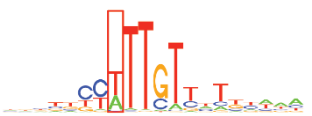
Table 1. Significant SNPs from antipsychotic pharmacogenomic GWAS identified by the NHGRI GWAS Catalog (cont.).

Study (year)	Sample		Variant	Locus	Gene	Response measurement	p-value	Effect†	Ref.
	Size: initial; replication	Ethnicity							
Liou *et al.* (2012)	522 cases and 806 controls; 273 cases	100% Han Chinese	rs230529	4q24	*NFKB1* (intron)	Treatment-refractory schizophrenia	2×10^{-7}§	+	[46]
			rs11265461	1q23.3	Intergenic		2×10^{-7}§	+	
			rs10218843	1q23.3	Intergenic		3×10^{-7}§	+	
Malhotra *et al.* (2012)	139 73; 40; 92	55% Cau, 23% AA, 22% other; 70% Cau, 30% AA; 100% Cau; 100% Cau	rs489693	18q21.32	Intergenic	Severe weight gain, several other metabolic indices	6×10^{-12}§	+	[47]
Clark *et al.* (2013)	738‡	57% EA, 29% AA, 14% other	rs8050896	16q22.1	Intergenic	CGI-S	4×10^{-8}	-	[6]
			rs17382202	5q12.1	*PDE4D* (intron)	PGI	4×10^{-8}	-	
			rs10170310	2q22.1	*SPOPL* (intron)	PGI	1×10^{-7}	+	
			rs6688363	1q23.2	Intergenic	CGI-S	2×10^{-7}	+	
			rs7395555	11q14.1	Intergenic	CGI-S	2×10^{-7}	-	
			rs17742120	5q12.1	*PDE4D* (intron)	PGI	2×10^{-7}	-	
			rs2164660	5q12.1	*PDE4D* (intron)	PGI	2×10^{-7}	-	
			rs711355	15q13.1	Intergenic	PGI	2×10^{-7}	-	
			rs2980976	18q21.3	Intergenic	CGI-S	3×10^{-7}	+	
			rs2636697	4q24	*PPA2* (intron)	CGI-S	4×10^{-7}	+	
			rs2636719	4q24	*PPA2* (intron)	CGI-S	5×10^{-7}	+	
			rs785423	15q13.1	Intergenic	PGI	5×10^{-7}	-	
			rs813676	15q13.1	Intergenic	PGI	5×10^{-7}	-	

†Direction of effect of minor allele, where '+' denotes minor allele frequency associated with poorer response or presence of ADRs.
‡Identical cohort from the CATIE study [40].
§Joint probability from meta-analysis of initial and replication cohorts.
AA: African–American; ADR: Adverse drug reaction; AIMS: Abnormal involuntary movement scale; Cau: Caucasian; CGI-S: Clinical global impression severity scale; EA: European–American; EPS: Extrapyramidal symptom; Hemoglobin A1c: Glycohemoglobin (used to measure plasma glucose levels); HDL: High-density lipoprotein; PANSS: Positive and negative syndrome scale; PGI: Patient global impression scale; QTc: Interval between ventricular depolarization (Q wave) and repolarization (T wave) in electrocardiogram, corrected for heart rate; SAS: Simpson–Angus scale.

10.2217/pgs-2016-0108

Table 2. Predicted expression quantitative trait loci from RegulomeDB with associated regulatory targets and effects.

SNP	Position (GRCh37/ hg19)	Score†	eQTL target(s)	Bound protein(s)	Sequence motif(s)		Other evidence	
					Name	PWM‡	Chromatin changes	Histone modifications
rs3774959	4:103511113	1b	*MANBA, COL9A2*	RFX3	Lmo2complex		Yes	Yes
rs230505	4:103481350	1d	*MANBA, COL9A2*	BATF	E12, Six4, Myf6, MyoD		Yes	Yes
rs230532	4:103450166	1f	*MANBA, COL9A2*	None	Nanog		Yes	Yes
rs230520	4:103465611	1f	*MANBA, COL9A2*	None	None		Yes	Yes
rs1599961	4:103443568	1f	*MANBA, COL9A2*	None	None		Yes	Yes
rs230504	4:103481560	1f	*MANBA, COL9A2*	None	None		Yes	Yes
rs230493	4:103486215	1f	*MANBA, COL9A2*	RFX3	Six-1		Yes	Yes
rs747559	4:103414174	1f	*MANBA, COL9A2*	None	Pitx2, Cdc5		Yes	Yes
rs4648055	4:103515312	1f	*MANBA, COL9A2*	FOS	None		Yes	Yes

†Score defined by Boyle *et al.* [11].
‡PWM corresponds to motif in bold. The red box indicates the SNP position.
PWM: Position weight matrix.

lines of evidence across different cell types that point to regulatory function in this region, including transcription factor binding sites, histone marks and open chromatin, and DNase I hypersensitive sites (Figure 2).

The rSNPbase tool predicted that 219 of the 531 SNPs affected regulation, either proximally, distally or post-transcriptionally. Notably, 104 of the 219 rSNPs were predicted to affect the expression of *NFKB1*. Furthermore, the nine eQTLs for *COL9A2* and *MANBA* that were identified by RegulomeDB were assigned the same function by rSNPBase and were also shown to affect *NFKB1*.

To further investigate the role of these variants on gene expression, the Genotype-Tissue Expression project database [49,50] was used for independent verification. These analyses revealed that the nine rSNPs predicted to be eQTLs for *COL9A2* and *MANBA* by RegulomeDB were also shown to significantly decrease the expression of *NFKB1* in adipose tissue (top $p = 4.3 \times 10^{-8}$ for rs230504 and rs230493), similar to the results obtained from rSNPBase. This highlights the important role that these variants are likely to play in regulating the expression of different genes.

Genetic association analyses

Genotyping was successful for all SNPs in Supplementary Table 1 and duplicate samples were concordant. All genotyped SNPs demonstrated a minor allele frequency ≥ 0.05 in the cohort of 103 patients, and are thus considered common variants within this cohort [51]. Additionally, all SNPs were in HWE ($p \geq 0.01$).

After Bonferroni adjustment for multiple testing, three variants were significantly associated with PANSS outcomes ($p < 0.0002$). Two of these variants (rs230504 and rs230493) were associated with poorer PANSS N outcomes, while the third SNP (rs3774959) was associated with poorer PANSS N and PANSS T outcomes. All three of these variants are in LD with a variant that was significantly associated with treatment-refractoriness by Liou *et al.* [46]. Table 3 shows these variants together with their associated outcomes, inheritance models and predicted effect sizes with 95% CIs.

All of these significantly associated variants were present in the 4q24 region. Further investigation of this region revealed that 14 SNPs in the 4q24 region (overlapping with the *NFKB1* transcription start site) are in strong LD within the FES cohort, similar to the LD observed in previous analysis of the Han Chinese cohort (Figure 3) [46]. The two designated haplotype blocks were analyzed further for associations with treatment outcomes and one of these haplotypes was significantly associated with worse positive symptom scores (Table 3). Additional analyses of haplotypes with frequencies greater than 0.01 are indicated in Supplementary Tables 3a–e.

In addition to those variants surpassing the Bonferroni adjusted significance threshold, there were ten SNPs and seven haplotypes that were associated with one or more treatment outcomes (unadjusted $p < 0.05$), with several of these variants associated with more than one symptom domain (Supplementary Table 4). PANSS Negative was the most significant and most fre-

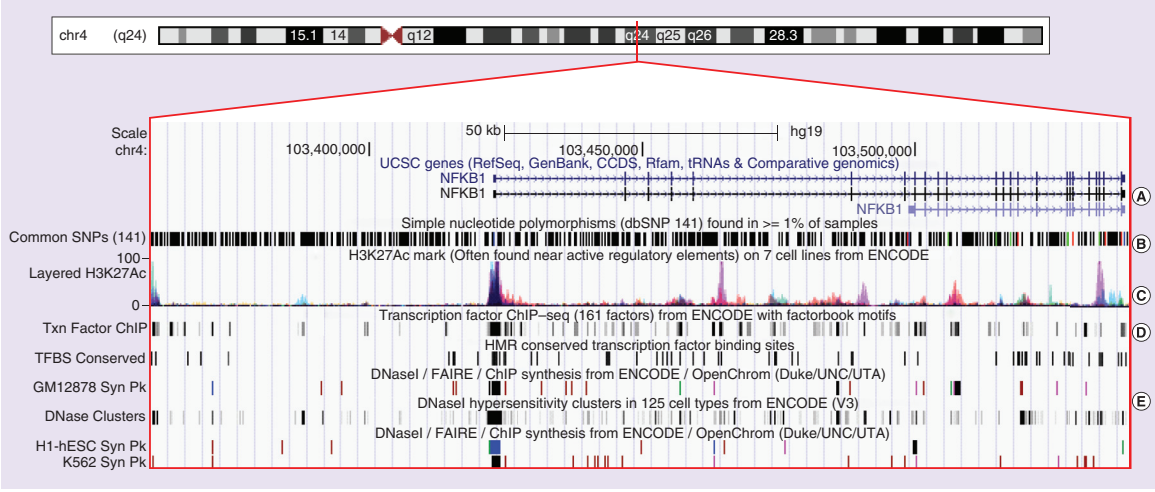


Figure 2. Magnified view of the 4q24 genomic region on the UCSC Genome Browser with ENCODE data tracks.
This region contains 14 of 19 predicted rSNPs according to RegulomeDB. ENCODE tracks show **(A)** transcription start of *NFKB1* gene, **(B)** common SNPs (minor allele frequency >1%) identified in this region, **(C)** peaks for histone mark H3K27Ac, associated with open chromatin, **(D)** TFBS by ChIP-Seq and **(E)** DNaseI hypersensitive sites determined by various experiments. Different colors indicate evidence in different cell lines.
TFBS: Transcription factor binding site.

Table 3. SNP and haplotype associations that survived correction for multiple testing ($p < 0.0002$).

Variant/haplotype	Associated response measurement	Inheritance model	Comparison	p-value	Effect†	95% CI
rs230504	PANSS negative	Dominant	TT+CT vs CC	1.48×10^{-4}	1.47	0.71–2.24
rs230493	PANSS negative	Genotype	TA vs TT	1.88×10^{-6}	1.98	1.20–2.76
			AA vs TT		0.38	-0.84–1.62
rs3774959	PANSS negative	Genotype	GA vs GG	1.75×10^{-5}	1.82	1.01–2.63
			AA vs GG		0.31	-0.85–1.48
	PANSS total	Genotype	GA vs GG	8.38×10^{-5}	1.67	0.91–2.44
			AA vs GG		0.72	-0.39–1.84
C.A.T.A.C.A.G.T.G.A.A.A.A.G‡	PANSS positive	Each additional haplotype		1.68×10^{-4}	4.41	2.10–6.78

†Effect measured in percentage change in PANSS score per month.
‡SNPs: rs230534, rs230532, rs230529, rs230526, rs118882, rs230520, rs230505, rs230504, rs230492, rs230493, rs230495, rs230539, rs3774959, rs4648055.
PANSS: Positive and Negative Syndrome Scale.

quently associated treatment outcome domain within the cohort.

Discussion

Previous GWAS designs

GWAS is a powerful tool that can be used to examine the genetics of antipsychotic treatment response outcomes, however, these data only point toward regions of interest and do not identify causal variants. Although the majority of antipsychotic GWAS performed to date acknowledge that the intergenic findings may affect regulatory processes, for the most part, the GWAS follow the trend of interpreting function in terms of the closest gene, and deprioritize genes that have not previously been implicated in antipsychotic response, neurological functioning or schizophrenia. All of the GWAS included in this study investigated LD to a limited extent, with some performing haplotype analyses, but once again these variants or haplotypes were related back to the closest gene. By implementing the strategy to rigorously investigate regulatory variation, this study has attempted to address these deficits in past research.

Significant rSNPs & functional implications

Bioinformatic analyses revealed that the results from both RegulomeDB and rSNPBase were enriched for the chromosome 4q24 region. Of the 19 SNPs predicted to be of regulatory importance by RegulomeDB, 14 occur at this locus. Nine of these variants received a score of 1, highlighting that variants in this region (4q24) have been experimentally linked to changes in the expression of specific genes. Furthermore, the ENCODE tracks in this region show extensive evidence of regulation in different cell lines. The specific region also includes the start of transcription of the *NFKB1* gene. Interestingly, the nine most significant SNPs were classified as eQTLs for *MANBA* (*cis*-acting) and *COL9A2* (*trans*-acting) by both RegulomeDB and rSNPBase [52]. This illustrates the importance of thoroughly assessing available regulatory information rather than assuming that a variant's effects are limited to the closest gene.

The eQTL evidence for this group of SNPs originates from a large-scale study by Zeller *et al.* [52] in which the transcriptome of monocytes was assessed in 1490 individuals. These nine variants are most likely reflective of one signal associated with the expression of these two genes. Given that this locus is associated with expression of these two genes, and is also associated with treatment response in our cohort, it is possible that there is a relationship between expression of these genes and treatment outcomes, however this requires further investigation. While neither gene has been previously associated with antipsychotic treatment response, both have shown links to schizophrenia. Variation in the *MANBA* gene is responsible for the development of β-mannosidosis, a lysosomal storage disorder [53]. These disorders are often accompanied by psychiatric disturbances including cognitive impairments, psychosis, schizophrenia and mood disorders [54]. Furthermore, a study by Altar *et al.* [55] on schizophrenia (independent of β-mannosidosis) found decreased expression of the *MANBA* gene in the hippocampal neurons of cases relative to controls. In a post-mortem analysis of gene expression in the dorsolateral prefrontal cortex of schizophrenia patients, levels of *COL9A2* mRNA were 12-fold lower in patients than controls [56]. While there are no other studies directly linking this gene to schizophrenia,

Glatt *et al.* [57] determined through analysis of blood mRNA that *COL9A2* was significantly dysregulated in children with autism, and there is substantial evidence for genetic overlap between different psychiatric disorders [58,59].

In addition to the evidence that was provided relating to the role that these variants play as eQTLs for *COL9A2* and *MANBA,* the rSNPBase tool predicted that the majority of input SNPs in the 4q24 region affected the expression of *NFKB1*. In fact, this gene was predicted to be affected by 104 SNPs, although no such evidence was obtained from RegulomeDB. The role that these variants play in *NFKB1* regulation was substantiated by Genotype-Tissue Expression analysis, which revealed that variants in this region are associated with decreased expression of the gene. *NFKB1* encodes part of a highly conserved transcription factor that regulates over 200 genes, and plays important roles in cancer and the immune system [46,60]. To our knowledge, there is no evidence that the NF-κB transcription factor interacts with either *MANBA* or *COL9A2*. Nonetheless, because NF-κB is a wide-acting transcription factor, it is possible that the gene expression results obtained by Zeller *et al.* [52] reflect the downstream effects of a decrease in *NFKB1* gene expression and NF-κB activity. This hypothesis requires validation in future studies.

There is a long-standing hypothesis that schizophrenia development is associated with abnormal immune functioning. In fact, the major histocompatibility locus is the most replicated genomic region with regard to associations with schizophrenia risk, and is a target of the NFκB transcription factor [61]. Several other immune factors were implicated in the regulatory analyses of this study, including RFX3 [62] and the transcription factors CEBP [63] and EBF1 [64]. These results suggest that regulation of different aspects of the immune system could contribute to variation in schizophrenia treatment response, particularly nonresponse. Supporting this idea, a recent

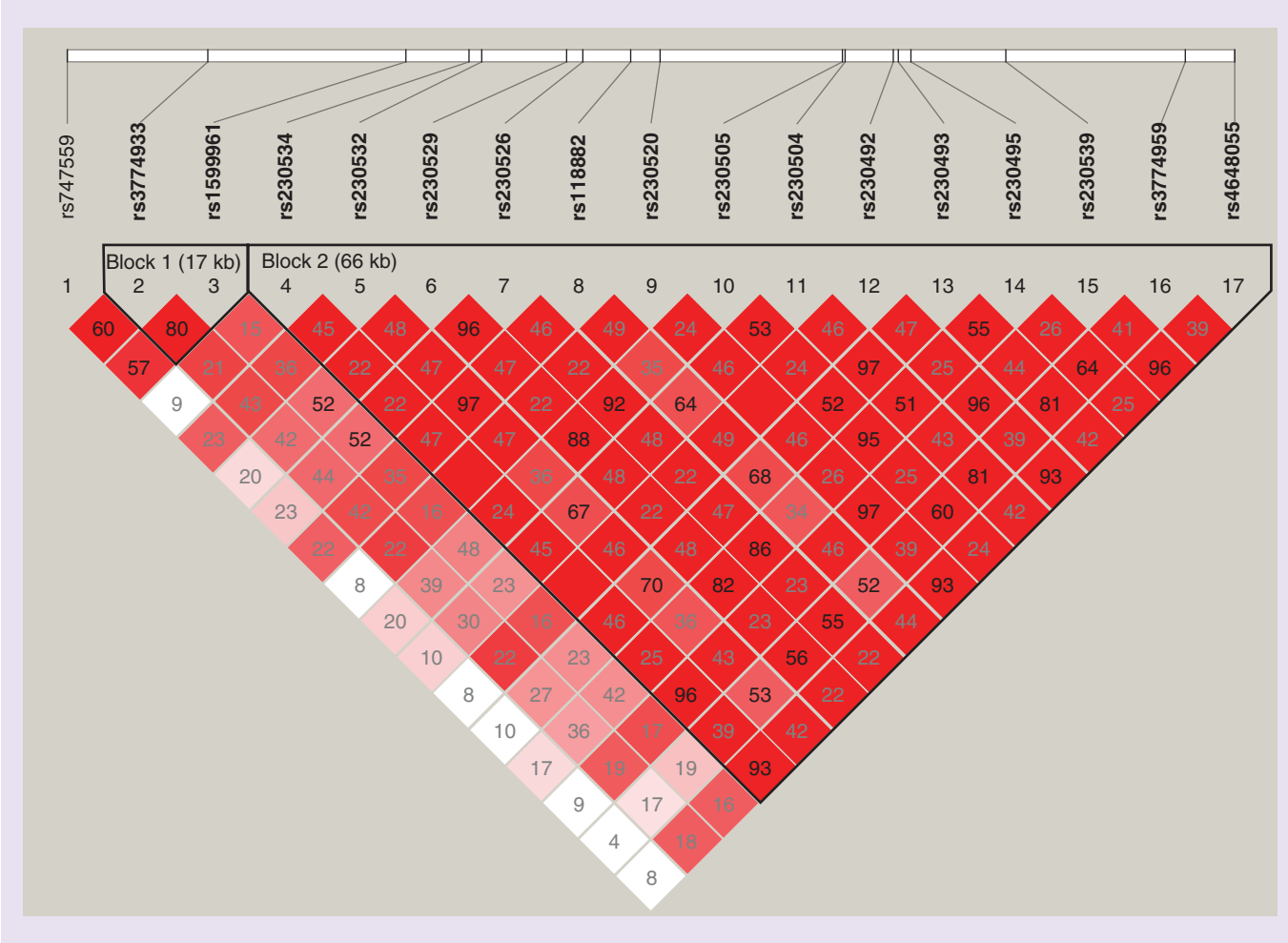


Figure 3. Two haplotype blocks on chromosome 4, designated by Haploview version 4.2 ($r^2 \geq 0.8$; LOD ≥3) [32]**.** Dark red squares indicate significant linkage disequilibrium between SNPs in the cohort; r^2 values are shown as a percentage in each square.

meta-analysis of 23 studies revealed that antipsychotics produce anti-inflammatory effects in schizophrenia [65]. Having said this, direct involvement of *NFKB1* and other immunity factors in antipsychotic treatment response remains to be determined.

Validation of findings through association analyses in the South African first episode schizophrenia cohort

The association analyses confirmed the importance of the 4q24 region in antipsychotic treatment response. Many SNPs in this region were nominally associated with PANSS outcomes, although only the associations of rs230504, rs230493 and rs3774959 with negative symptoms, and rs3774959 with total symptoms survived Bonferroni correction (Table 3). None of these variants have previously been associated with treatment outcomes, however the initial GWAS variant (rs230529) that is in LD with these SNPs was associated with treatment refractoriness [46], providing additional evidence for the role that this region plays in treatment outcomes. These results illustrate the limitations of tag SNPs utilized in GWAS. All of the 4q24 rSNPs were identified from LD analysis of rs230529, however, this original GWAS variant was not predicted by RegulomeDB to affect regulatory regions significantly, even though all the SNPs in this region were in strong LD for three independent population groups. Further substantiating this, the tag SNP was not significantly associated with any treatment outcomes in the South African cohort, despite having a relatively high frequency of 0.46. The lack of associations for rs230529 reveals that this SNP is not an accurate proxy for the region in all populations and illustrates the importance of analyzing each variant at a locus to identify the causal variant.

In addition to pinpointing potentially causal variants, the association of the variants identified in this study with negative symptomology in our cohort is encouraging. The negative symptoms of schizophrenia, in other words, avolition and blunted emotion, are especially complex and difficult to treat [66]. These symptoms tend to persist even when positive symptoms have improved, and have been shown to influence the extent of residual cognitive deficits and functional outcomes in schizophrenia patients [67–69]. Therefore, these results could lead to further insight into the mechanisms of negative symptom treatment. Furthermore, the top associations within the FES cohort were all predicted to be associated with less improvement in PANSS score outcomes after antipsychotic treatment, with effect sizes ranging from 0.31 to 1.98% per month for individual SNPs. This relative lack of improvement points to nonresponse for these symptom domains, which is consistent with the association observed by Liou *et al.* [46], and likely contributes to treatment-refractoriness. These three SNPs also have eQTL evidence linking them to *MANBA*, *COL9A2* and *NFKB1* [52], suggesting that aberrant expression could contribute to poor outcomes in the negative symptom domain.

The large haplotype on chromosome 4 that contains the minor alleles for rs230504, rs230493 and rs3774959 was nominally associated with PANSS positive, negative and total scores, and maintained significance with positive scores after correction for multiple testing (Table 3, Supplementary Table 4). The three individual variants were also nominally associated with positive symptom outcomes. These results suggest that, in combination, the three significant alleles in the haplotype produce an additive effect in relation to positive symptom changes over the course of treatment. Although this haplotype occurred at a low frequency in the cohort (0.03), its presence is associated with a substantial difference in PANSS positive scores (4.41% per month higher). Therefore, these variants and different haplotypic combinations should be further studied in larger cohorts to validate these findings.

Study limitations & future directions

Limitations of our study include the fact that the ENCODE data are relatively recent, and the bioinformatics tools that have stemmed from its release are still in their infancy. The limited consensus between the tools demonstrates the need for improved bioinformatic design and database curation, as well as the complexity of genetic regulation. Second, ENCODE has received criticism for its claim that 80% of the genome is functional [70], and one must keep in mind that the data may overestimate functionality.

Importantly, this study is limited by the comparison of significant variants between different treatment cohorts and outcomes from previous GWAS. There are several factors that restrict direct comparison, and thus the results should be interpreted with caution and validated in other cohorts. Our ability to only replicate significant associations with variants in regions previously associated with efficacy may relate to the fact that genetic variants are not transferable across different treatment phenotypes (e.g., weight gain).

While previous GWAS studies cumulatively made use of a variety of first and second generation antipsychotics, the South African cohort was treated with a first generation antipsychotic due to its low cost and availability. Although we acknowledge that our findings may not be generalizable to patients treated with other agents, we consider this unlikely for the follow-

ing reasons: first, first generation and second generation antipsychotics are not homogeneous classes; in fact it has been recommended that this distinction be abandoned [71]. Second, flupenthixol is regarded as a 'partially atypical' antipsychotic [72], sharing receptor binding characteristics of several of the second generation antipsychotics [73]. There are several studies providing evidence that genetic variants associated with treatment response are transferable across antipsychotic classes [39,40,74]. In any event, the variants that *were* significantly associated with treatment outcomes in this study (directly or by LD proxy) may point to general or overlapping mechanisms between different antipsychotic classes. Future studies examining replicative, drug-specific cohorts and different treatment phenotypes are warranted.

The FES cohort is small compared with, for example, the CATIE cohort. However, it is important to note that the FES cohort is extremely well-characterized and homogenized. Considering the clinical stringency and limited confounders, this study provides increased statistical power [75,76]. The importance of this has been well-documented in a study by Malhotra *et al.* [47], which demonstrated the ability to detect and replicate a genome-wide signal in a patient group of comparable size (n = 139) to the FES cohort. Our study did not have a validation cohort available to confirm associations, however, as the regions of interest investigated were identified via past antipsychotic GWAS, this study may be considered a replication of these findings, the results of which have been further substantiated by bioinformatic predictions.

Although this study has revealed potential candidate genes for antipsychotic response, there are several aspects of the design that require further research and functional validation. First, although the LD-inclusive and fine-mapping approach of this study has aided in narrowing down the signals responsible for significant associations, further study is required before presuming causal relationships between, for example, the chromosome 4q24 locus and treatment outcomes. Additionally, it remains to be determined which regulatory mechanism(s) or gene product(s) are responsible for alterations in treatment response. Each of the three genes hypothetically connected to treatment response should be tested in, for example, a knock-out model combined with antipsychotic treatment, to determine their functional consequences. This future research will be aided by tissue-specific culture or animal model work in order to determine if these factors are expressed in tissues relevant to schizophrenia and drug response, such as the brain and the liver.

Conclusion

This study successfully used a unique bioinformatics approach validated by genetic association analyses to identify variants associated with schizophrenia treatment response, and can potentially be applied to other disorders. The all-encompassing nature of this approach – in other words, analysis of LD regions, focus on noncoding variants and evaluation of diverse, well-characterized clinical outcomes – has led to the formation of new hypotheses regarding the biology of antipsychotic mechanisms and treatment response. The suggested relationship between 4q24 variation, *COL9A2*, *MANBA*, *NFKB1* and negative symptomology remains undefined, but is a promising avenue of investigation for further study, given that negative symptoms are poorly understood and difficult to treat. This study has identified potential causal variation by demonstrating enrichment of the 4q24 region in regulation and downstream affected genes that have established links to schizophrenia. The connections to regulatory functioning, immunity and treatment nonresponse suggest that this locus has important and widespread implications in schizophrenia, and should be further investigated.

Executive summary

- Despite the promise of GWAS, there is little replication and functional validation of the plethora of results generated, particularly with regard to noncoding variants.
- The ENCODE results, coupled with user-friendly online prediction tools, enable thorough investigation of the role of noncoding variants in regulatory mechanisms.
- Investigation of previous GWAS results revealed enrichment for regulatory variants potentially affecting *NFKB1*, *MANBA* and *COL9A2* expression.
- Three variants and one haplotype in this region were associated with poor treatment outcomes in both positive and negative symptom domains.
- These findings support the association of variation in this region with treatment-refractoriness in a group of Han Chinese individuals [46].
- Discovering associations with negative symptoms is encouraging, since this symptom domain is characteristically difficult to treat.
- Suggested associations between *COL9A2*, *MANBA*, *NFKB1* and antipsychotic treatment response were uncovered, and this relationship should be investigated further through functional validation.

Supplementary data

To view the supplementary data that accompany this paper, please visit the journal website at: www.futuremedicine.com/doi/full/10.2217/pgs-2016-0108

Acknowledgements

The authors acknowledge M Coffee for her contribution to the development of the bioinformatics approach.

Financial & competing interests disclosure

The authors acknowledge the following funding sources for this study: The National Research Foundation of South Africa; The Medical Research Council of South Africa; New Africa Partnership Developments (NEPAD) through the Department of Science and Technology, South Africa; and the L'Oréal-UNESCO for women in Science in Sub-Saharan Africa Fellowship. The authors have no other relevant affiliations or financial involvement with any organization 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 utilized in the production of this manuscript

Ethical conduct of research

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

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest

1 Blanc O, Brousse G, Meary A, Leboyer M, Llorca P-M. Pharmacogenetics of response efficacy to antipsychotics in schizophrenia: pharmacodynamic aspects. Review and implications for clinical research. *Fundam. Clin. Pharmacol.* 24, 139–160 (2010).

2 Klein K, Zanger U. Pharmacogenomics of cytochrome P450 3A4: recent progress toward the "missing heritability" problem. *Front. Genet.* 4(12), 1–15 (2013).

3 Ni X, Zhang W, Huang RS. Pharmacogenomics discovery and implementation in genome-wide association studies era. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 5, 1–9 (2013).

4 Arranz MJ, de Leon J. Pharmacogenetics and pharmacogenomics of schizophrenia: a review of last decade of research. *Mol. Psychiatry* 12, 707–747 (2007).

5 Åberg K, Adkins DE, Liu Y, McClay JL, Bukszár J, Jia P. Genome-wide association study of antipsychotic induced QTc interval prolongation. *Pharmacogenomics* 12(2), 165–172 (2012).

6 Clark SL, Souza RP, Adkins DE *et al.* Genome-wide association study of patient-rated and clinician-rated global impression of severity during antipsychotic treatment. *Pharmacogenet. Genomics* 23(2), 69–77 (2013).

7 Cooper G, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat. Rev. Genet.* 12(9), 628–640 (2011).

8 The ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 9(4), e1001046 (2012).

• **An excellent summary of the ENCODE project and its results.**

9 Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res.* 22, 1748–1759 (2012).

10 Maurano MT, Humbert R, Rynes E *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195 (2013).

•• **Study stemming from the ENCODE results that illustrates enrichment for disease variants in noncoding regions.**

11 Boyle AP, Hong EL, Hariharan M *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22(9), 1790–1797 (2012).

12 Guo L, Du Y, Chang S, Zhang K, Wang J. rSNPBase: a database for curated regulatory SNPs. *Nucl. Acids Res.* 42(database issue), D1033–D1039 (2014). http://nar.oxfordjournals.org/

13 Welter D, MacArthur J, Morales J *et al.* The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42(D1), D1001–D1006 (2013).

14 NHGRI GWAS Catalog. www.genome.gov/gwastudies/

15 Kingsmore SF, Lindquist IE, Mudge J, Gessler DD, Beavis WD. Genome-wide association studies: progress and potential for drug discovery and development. *Nat. Rev. Drug Discov.* 7(3), 221–230 (2008).

16 Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.* 74, 106–120 (2004).

17 SNAP version 2.2. www.broadinstitute.org/mpg/snap/

18 The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 468, 1061–1073 (2010).

19 The International HapMap Consortium. The international HapMap project. *Nature* 426, 789–796 (2003).

20 RegulomeDB. http://regulome.stanford.edu/

21 Chung CC, Kanetsky PA, Wang Z *et al.* Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nat. Genet.* 45(6), 680–685 (2013).

22 rSNPBase. http://rsnp.psych.ac.cn/

23 Chiliza B, Asmal L, Kilian S, Phahladira L, Emsley R. Rate and predictors of non-response to first-line antipsychotic

treatment in first-episode schizophrenia. *Hum. Psychopharm. Clin.* 30(3), 173–182 (2015).

• **Further clinical information on the cohort used in this study.**

24 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-I). 1st Edition* American Psychiatric Association, Washington DC, USA (1994).

25 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, 441–449 (2005).

26 Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* 13(2), 261–276 (1987).

27 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).

28 Drögemöller BI, Niehaus DJH, Chiliza B *et al.* Patterns of variation influencing antipsychotic treatment outcomes in South African first-episode schizophrenia patients. *Pharmacogenomics* 15(2), 189–199 (2014).

29 SNP Genotyping Assay Search Tool. www.lifetechnologies.com/

30 Solé X, Guinó E, Valls J, Iniesta R, Moreno V. SNPStats: a web tool for the analysis of association studies. *Bioinformatics* 22(15), 1928–1929 (2006).

31 SNPStats. http://bioinfo.iconcologia.net/snpstats/start.htm

32 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2), 263–265 (2005).

33 Sinnwell JP, Schaid DJ. haplo.stats: statistical analysis of haplotypes with traits and covariates when linkage phase is ambiguous. R package version 1.7.1 (2015). http://CRAN.R-project.org/package=haplo.stats

34 Pinheiro J, Bates D, DebRoy S *et al.* nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1–117 (2014). http://CRAN.R-project.org/package=nlme

35 Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67(1), 1–48 (2015).

36 Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0–32 (2016). https://CRAN.R-project.org/package=lmerTest

37 Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9), 1655–1664 (2010).

38 Daya M, van der Merwe L, Galal U *et al.* A panel of ancestry informative markers for the complex five-way admixed South African coloured population. *PloS ONE* 8(12), e82224 (2013).

39 Drögemöller B, Emsley R, Chiliza B *et al.* The identification of novel genetic variants associated with antipsychotic treatment response outcomes in first episode schizophrenia patients. *Pharmacogenet. Genomics* 26(5), 235–242 (2016).

40 Lieberman JA, Stroup S, McEvoy J *et al.* Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N. Engl. J. Med.* 353, 1209–1223 (2005).

41 Åberg K, Adkins DE, Bukszár BT *et al.* Genomewide association study of movement-related adverse antipsychotic effects. *Biol. Psychiatry* 67(3), 279–282 (2010).

42 Adkins DE, Åberg K, McClay JL *et al.* Genomewide pharmacogenomic study of metabolic side effects to antipsychotic drugs. *Mol. Psychiatry* 16(3), 321–322 (2011).

43 McClay JL, Adkins DE, Åberg K *et al.* Genome-wide pharmacogenomics study of neurocognition as an indicator of antipsychotic treatment response in schizophrenia. *Neuropsychopharmacology* 36(3), 616–626 (2011).

44 McClay JL, Adkins DE, Åberg K *et al.* Genome-wide pharmacogenomic analysis of response to treatment with antipsychotics. *Mol. Psychiatry* 16, 76–85 (2011).

45 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. *Psychiatry Res.* 197(3), 327–336 (2012).

46 Liou Y-J, Wang H-H, Lee M-TM *et al.* Genome-wide association study of treatment refractory schizophrenia in Han Chinese. *PloS ONE* 7(3), e33598 (2012).

•• **Study demonstrating the link between *NFKB1* and treatment refractoriness.**

47 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).

•• **Example of an extremely well-characterized cohort in an antipsychotic response genome-wide association studies.**

48 ENCODE: UCSC Genome Browser. http://genome.ucsc.edu/ENCODE/

49 The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348(6235), 648–660 (2015).

50 GTEx. http://www.gtexportal.org

51 Iyegbe C, Campbell D, Butler A, Ajnakina O, Sham P. The emerging molecular architecture of schizophrenia, polygenic risk scores and the clinical implications for GxE research. *Soc. Psychiatry Psychiatr. Epidemiol.* 49(2), 169–182 (2014).

52 Zeller T, Wild P, Szymczak S *et al.* Genetics and beyond – the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* 5(5), e10693 (2010).

53 Levade T, Graber D, Flurin V *et al.* Human beta-mannosidase deficiency associated with peripheral neuropathy. *Ann. Neurol.* 35(1), 116–119 (1994).

54 Staretz-Chacham O, Hyuk Choi J, Wakabayashi K *et al.* Psychiatric and behavioural manifestations of lysosomal storage disorders. *Am. J. Med. Genet. Part B* 153B, 1253–1265 (2010).

55 Altar CA, Jurata LW, Charles V *et al.* Deficient hippocampal neuron expression of proteasome, ubiquitin, and mitochondrial

genes in multiple schizophrenia cohorts. *Biol. Psychiatry* 58, 85–96 (2005).

56 Dean B, Keriakous D, Scarr E, Thomas EA. Gene expression profiling in Brodmann's area 46 from subjects with schizophrenia. *Aust. N. Z. J. Psychiatry* 41, 308–320 (2007).

57 Glatt SJ, Truang MT, Winn M *et al.* Blood-based gene expression signatures of infants and toddlers with autism. *J. Am. Acad. Child Adolesc. Psychiatry* 51(9), 934.e2–944.e2 (2012).

58 Adam D. On the spectrum. *Nature* 496(7446), 416–418 (2013).

59 Doherty JL, Owen MJ. Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. *Genome Med.* 6(4), 29–41 (2014).

60 Shishodia S, Aggarwal BB. Nuclear factor-κB activation mediates cellular transformation, proliferation, invasion angiogenesis and metastasis of cancer. *Cancer Treat. Res.* 119, 139–173 (2004).

61 Sullivan PF, Daly M, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.* 13(8), 537–551 (2012).

62 Reith W, Steimle V, Mach B. Molecular defects in the bare lymphocyte syndrome and regulation of MHC class II genes. *Immunol. Today* 16(11), 539–546 (1995).

63 Marigo I, Bosio E, Solito S *et al.* Tumor-induced tolerance and immune suppression depend on the C/EBPβ transcription factor. *Immunity* 32(6), 790–802 (2010).

64 Nechanitzky R, Akbas D, Scherer S *et al.* Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. *Nature Immunol.* 14, 867–875 (2013).

65 Tourjman V, Kouassi É, Koué M-É *et al.* Antipsychotics' effects on blood levels of cytokines in schizophrenia: a meta-analysis. *Schizophr. Res.* 151(1–3), 43–47 (2013).

66 Millan MJ, Fone K, Steckler T, Horan WP. Negative Symptoms of schizophrenia: clinical characteristics, pathophysiological substrates, experimental models and prospects for improved treatment. *Eur. Neuropsychopharmacol.* 24(5), 645–692 (2014).

67 Lin C-H, Huang C-L, Chang Y-C *et al.* Clinical symptoms, mainly negative symptoms, mediate the influence of neurocognition and social cognition on functional outcome of schizophrenia. *Schizophr. Res.* 146, 231–237 (2013).

• **More information on the complexity of negative symptoms.**

68 Malaspina D, Walsh-Messinger J, Gaebel W *et al.* Negative symptoms, past and present: a historical perspective and moving to DSM-5. *Eur. Neuropsychopharmacol.* 24(5), 710–724 (2014).

69 Woodward TS, Jung K, Smith GN *et al.* Symptom changes in five dimensions of the Positive and Negative Syndrome Scale in refractory psychosis. *Eur. Arch. Psychiatry Clin. Neurosci.* 264(8), 673–682 (2014).

70 Graur D, Zheng Y, Price N, Azevedo RBR, Zufall RA, Elhaik E. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol. Evol.* 5(3), 578–590 (2013).

71 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, 31–41 (2009).

72 Gattaz WF, Diehl A, Geuppert MS *et al.* Olanzapine versus flupenthixol in the treatment of inpatients with schizophrenia: a randomized double-blind trial. *Pharmacopsychiatry* 37, 279–285 (2004).

73 de Wit H. Flupenthixol. In: *Encyclopedia of Psychopharmacology.* Stolerman IP (Ed.). Springer, Berlin, Germany, 538–539 (2010).

74 Ramsey TL, Brennan MD. Glucagon-like peptide 1 receptor (*GLP1R*) haplotypes correlate with altered response to multiple antipsychotics in the CATIE trial. *Schizophr. Res.* 160(1–3), 73–79 (2014).

75 Reynolds GP. The impact of pharmacogenetics on the development and use of antipsychotic drugs. *Drug Discov. Today* 12(21–22), 953–959 (2007).

76 Zhang J-P, Malhotra AK. Pharmacogenetics of antipsychotics: recent progress and methodological issues. *Expert Opin. Drug Metab. Toxicol.* 9(2), 183–191 (2013).

• **Excellent summary of research in the field of antipsychotic pharmacogenomics.**