Pardiñas, A. F., Nalmpanti, M., Pocklington, A., Legge, S., Medway, C., King, A., Jansen, J., Helthuis, M., Zammit, S., MacCabe, J. H., Owen, M. J., O'Donovan, M. C., & Walters, J. T. (2019). Pharmacogenomic variants and drug interactions identified through the genetic analysis of clozapine metabolism. *American Journal of Psychiatry*, *176*(6), 477-486. https://doi.org/10.1176/appi.ajp.2019.18050589

Number of words (abstract + text + references): 6,299

Number of tables: 3; Number of figures: 2

Pharmacogenomic variants and drug interactions identified through the genetic analysis of clozapine metabolism

Antonio F. Pardiñas[1], Ph.D., Mariana Nalmpanti[1], B.Sc., Andrew J. Pocklington[1], Ph.D., Sophie E. Legge[1], Ph.D., Christopher Medway[1], Ph.D., Adrian King[2], Ph.D., John Jansen[3], Ph.D., Marinka Helthuis[3], M.D., Stanley Zammit[1,4,5],MRCPsych., Ph.D., James MacCabe[6], FRCPsych., Ph.D., Michael J. Owen[1], FRCPsych., Ph.D., Michael C. O'Donovan[1], FRCPsych., Ph.D., James T. R. Walters[1,*] , MRCPsych., Ph.D.

[1] MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK

[2] Magna Laboratories Ltd., Ross-on-Wye, UK

[3] Leyden Delta B.V., Nijmegen, Netherlands

[4] Centre for Academic Mental Health, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[5] National Institute for Health Research (NIHR) Biomedical Research Centre, University of Bristol, Bristol, United Kingdom

[6] Department of Psychosis Studies, Institute of Psychiatry Psychology and Neuroscience, King's College London, London, UK

* Corresponding author

MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University

Hadyn Ellis Building, Maindy Road, Cardiff, CF24 4HQ, UK

E-mail: WaltersJT@cardiff.ac.uk

Phone: +44 (0)29 20688 434

Acknowledgements

This project was supported by grants from the Medical Research Council (MRC) to Cardiff University (MR/L010305/1; G0800509; MR/L011794/1 and MC_PC_17212). The project has also received funding from the European Union's Seventh Framework Programme (279227). This publication reflects only the authors' views and the European Union is not liable for any use that may be made of the information contained therein. J.M. is partly funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London.

We acknowledge Leyden Delta and Magna Laboratories for supporting the CLOZUK2 sample collection, anonymisation and data preparation (Andy Walker and Anouschka Colson). We acknowledge deCODE genetics (Hreinn Stefansson) for genotyping of the CLOZUK2 sample. We acknowledge Cardiff University MRC Centre Core Team (Lucinda Hopkins, Lesley Bates, and Catherine Bresner) for laboratory sample management. We acknowledge Cardiff University MRC Centre HPC team (Mark Einon), and Cardiff University Advanced Research Computing division (Wayne Lawrence), for support with the use and setup of computational infrastructures.

Competing Financial Interests

M. H. and J. J. are full-time employees of Leyden Delta B.V.; A.K. is a full-time employee of Magna Laboratories Ltd. The remaining authors declare no conflicts of interest.

Abstract

Objective: Clozapine is the only effective medication for treatment-resistant schizophrenia (TRS), but its worldwide use is still very limited due to its complex titration protocols. While the discovery of pharmacogenomic variants of clozapine metabolism might improve clinical management, no robust findings have yet been reported. Our study is the first to adopt the framework of genome-wide association studies to discover genetic markers of clozapine plasma concentrations in a large sample of TRS patients.

Method: We used mixed-model regression to combine data from multiple assays of clozapine metabolite plasma concentrations from a clozapine monitoring service, and carried out a genome-wide analysis of clozapine, norclozapine and their ratio on 10,353 assays from 2,989 individuals. We adjusted these analyses for demographic factors known to influence clozapine metabolism, although it was not possible to adjust for all potential mediators given the available data. GWAS results were used to pinpoint specific enzymes and metabolic pathways, and compounds which might interact with clozapine pharmacokinetics.

Results: We identify four distinct genome-wide significant loci, which harbour common variants impacting the metabolism of clozapine or its metabolites. Detailed examination pointed to coding and regulatory variants at several CYP* and UGT* genes, and corroborative evidence for interactions between the metabolism of clozapine, coffee and tobacco. Individual effect of single SNPs fine-mapped from these loci were large, such as the minor allele of rs2472297, which was associated with a reduction in clozapine concentrations roughly equivalent to a decrease in clozapine dose of 50 mg/day. On their own, these single SNPs explained from 1.15% to 9.48% of the variance in our plasma concentration data.

Conclusions: Common genetic variants with large effects on clozapine metabolism exist and can be found via genome-wide approaches. Their identification opens the way for clinical studies assessing the use of pharmacogenomics in the clinical management of TRS patients.

Introduction

Schizophrenia affects approximately 0.7% of the population (1), is characterized by disturbances in cognition, emotion, perception and thought, and severely impacts quality and length of life (2). Around a third of patients experience treatment-resistant schizophrenia (TRS), a form of the disorder marked by severe functional impairment in which symptoms fail to respond adequately to at least two first line antipsychotics (3). Clozapine is the most effective treatment (4) and the only licensed medication for TRS. Despite extensive evidence supporting its effectiveness clozapine remains under-prescribed worldwide, including in countries with highly-developed health services (5, 6). A key factor limiting clozapine's use is its potential to induce severe adverse drug reactions (ADRs) including agranulocytosis, which occurs in up to 1% of patients and necessitates regular haematological monitoring (7). Other ADRs such as seizures, tachycardia, sedation, weight-gain, and hypersalivation have been associated with either clozapine dosage or high plasma concentrations (8).

Clinicians routinely use clozapine levels to assess adherence and guide dosage in the management of both therapeutic response and side effects (9). This is an important strategy as adverse effects are the primary reason for clozapine discontinuation (10). However, there is high inter-individual variability in plasma clozapine concentrations at given doses (11), also due to the effects of concomitant medication (12), which complicates titration and presents challenges for any research that aims to assess the relationship between dose, efficacy and ADRs. Guidelines for therapeutic drug monitoring (TDM) indicate that clozapine plasma concentrations in the range 0.35-0.60mg/L are optimal for response (13), while concentrations higher than 0.60mg/L have been linked to serious ADRs. Dose-response relationships between clozapine concentration and weight gain (14) or sedation (15) have been suggested, although this has not been seen for all ADRs (8). The accurate prediction of clozapine plasma levels therefore has important clinical implications. Sophisticated models incorporating lifestyle habits and metabolic indicators can explain up to 48% of the variance in clozapine levels in large patient samples (16, 17), but no individual factors other than age, smoking habits or sex have been found to be of clinical value (11).

The use of genetic approaches to identify biomarkers that explain individual variability in drug metabolism and response is the basis of the field of pharmacogenomics. While this discipline experienced a strong growth in the mid-2000s fostered by successes related to cardiology and oncology (18), translation of these results into clinical settings has been proven challenging (19).

This has been particularly true in psychiatry (20), and clozapine research in this area showcases some of the difficulties of carrying out robust pharmacogenomics studies. Clozapine is metabolised by the liver (21) with first-pass metabolism, driven primarily by the CYP1A2 enzyme, producing norclozapine (N-desmethylclozapine), a pharmacologically active compound that can reach up to 90% of the circulating concentration of clozapine (22, 23). Other metabolites have been identified, such as clozapine-N-oxide, formed by CYP3A4 (22), and N-glucuronides, which are secondary and tertiary metabolites produced by the UDP-glucuronosyltransferase (UGT) protein superfamily (24). These enzymes are all implicated in other drug metabolic pathways (25, 26), and have been the focus of the search for genetic variants associated with clozapine plasma concentrations. Previous studies have examined candidate polymorphisms in small (usually N<100) samples (27, 28). Promising results have been reported for variants of the ABCB1 drug transporter (28) but no finding has yet passed the threshold of genome-wide significance, which is now widely accepted as being required for robust association, even in candidate gene studies (29).

Here we report the first genome-wide association study (GWAS) of plasma concentrations of clozapine and its metabolites; by applying modern statistical modelling techniques, we exploit information from over 10,000 metabolite concentration assays taken from a sample of nearly 3,000 TRS patients. We identify genome-wide significant polymorphisms that delineate clozapine's metabolic pathways, and discuss their relevance to the clinical management of TRS.

Methods

Sample

Data were acquired as part of the CLOZUK2 study (30) from individuals prescribed clozapine for TRS in the UK. Sample and data acquisition were arranged through collaboration with Leyden Delta (Nijmegen, Netherlands), who monitor clozapine in the UK. The study was conducted in accordance with its UK NHS ethics permissions. The CLOZUK2 sample is described fully elsewhere (30).

Genotyping and imputation

Genotyping of the CLOZUK2 sample was performed by deCODE Genetics (Reykjavik, Iceland), using an Illumina HumanOmniExpress-12 array. Quality control (QC) and analyses were performed using PLINK v1.9 (31) unless otherwise specified. QC followed standard GWAS

protocols (32), including the removal of samples and markers with >2% missingness or homozygosity (F) >0.2. Post-QC, 7,287 individuals with data from 698,442 markers remained in the dataset.

The Haplotype Reference Consortium (HRC) panel, accessed through the Michigan Imputation Server (33, 34), was used for genotype imputation. Since using this service to impute X-chromosome data was not possible at the time of this study, genotype data from the X-chromosome was imputed locally using the Cardiff University RAVEN cluster (35). For this, the SHAPEIT/IMPUTE2 algorithms (36) and a combination of the 1000 Genomes phase 3 (1KGPp3) and UK10K reference panels (37) were used. Both approaches have been shown to perform similarly when imputing GWAS variants (33), which traditionally have minor allele frequencies (MAF) >1%. After imputation, 20 million SNPs with INFO scores >0.8 remained in the dataset.

Selection of individuals for analysis

In order to select a sample of individuals with homogeneous genetic ancestry, we selected a custom panel of ancestry informative markers (AIMs), as described previously (38). Briefly, using principal components and a classification algorithm based on linear discriminant analysis, we identified 5,900 CLOZUK2 individuals with >90% probability of European ancestry. Individuals not fulfilling this criterion were excluded from all further analysis, given the small number of non-European individuals with plasma concentration data.

From the imputed genotypes we retained SNPs with MAF ≥ 1% and Hardy-Weinberg Equilibrium (HWE) p-value ≤ 10^{-6}, leaving 7.5 million SNPs for analysis.

Clozapine and norclozapine levels in blood

Clozapine and norclozapine plasma concentration assays were conducted at Magma laboratories and were determined by Liquid Chromatography Mass Spectrometry (LC-MS/MS) using standard procedures for clozapine assays, summarised in **Supplementary Methods**.

Curation of plasma concentration data

The clozapine and norclozapine plasma concentrations of the CLOZUK2 sample formed a dataset of assays taken at 15,504 time points from 3,986 unique individuals. Data available for each time point included: age of patient; daily clozapine dose; date and time of last dose; date and time the blood was sampled; and measured clozapine and norclozapine levels. Clozapine or norclozapine concentrations <0.05 mg/L (corresponding to the minimal detection thresholds of the HPLC instrument) were removed as indicating non-adherence. We also excluded outliers outside the

99th percentile of plasma concentrations. We noted that removing data from a broader range of the extremes of the plasma concentration distributions did not meaningfully alter our results. Finally, we removed assays where the blood was sampled <6 hours or >24 hours since the last clozapine dose taken, since a 6-24 hour post-dose measurement interval is recommended for clozapine monitoring to ensure adequate drug absorption (11). After this process, 10,353 assays from 2,989 individuals (range of 1-42 per individual) remained (**Supplementary Figure 1**). 2,022 individuals (67.8% of the total) had assay data from more than 1 time point. We did not undertake analyses specifically on steady-state levels (39) given we could not strictly determine this on the available data, but noted that in restricting to those with more than one clozapine measurement (and hence more likely to have taken the medication for a longer period) the results were unchanged; indeed the data from the majority of our patients with >1 assay measurement spanned a much longer period (median time= 361 days).

Generation of plasma concentration phenotypes

We examined three primary metabolic outcome variables: the plasma concentrations of clozapine and norclozapine, as well as the clozapine-to-norclozapine ratio (the "metabolic ratio") (16). In order to make maximum use of the data available at multiple time points we employed a regression modelling framework to combine data from multiple assays into a single phenotype per individual, as detailed in **Supplementary Methods**. Briefly, we identified the best-fitting distribution for each metabolic outcome variable and used this to specify a random-effects model controlling for known predictors of clozapine metabolism (clozapine dose, time between dose and assay, and age at assay). From this model, random-effects coefficients were extracted for each individual; this corresponds to the variation in plasma concentrations for that individual (clozapine, norclozapine, and their ratio), independent of the effects of the known predictors (dose, time since dose and age); this co-efficient was then used as our primary outcome phenotype for our GWAS.

GWAS of plasma concentrations

The above approach to derive our primary outcome allowed us to use standard methods to perform a GWAS of clozapine concentration, norclozapine concentration and their metabolic ratio. We undertook the GWAS on the CLOZUK2 imputed data, using the "leave-one-chromosome-out" linear mixed-model (LMM-LOCO) implemented in GCTA v1.26 (40). This analysis requires genotype relatedness matrices to control for family and population structure, which we calculated from non-imputed genotypes to avoid introducing biases due to imputation

accuracy. Sex was used as a covariate for this analysis. As we did not have information on other factors known to influence drug plasma concentrations, such as weight or cigarette smoking habits, we performed secondary sensitivity analyses controlling for proxy measures based on polygenic risk scores (PRS) for those traits (**Supplementary Methods**). From each LMM-LOCO analysis we identified approximately independent index SNPs (r^2=0.1) using the PLINK linkage disequilibrium (LD) clumping procedure ($P < 10^{-4}$ and distance < 3000 kb).

Identification of putatively causal SNPs and genes

For each genome-wide significant locus ($P < 5x10^{-8}$), FINEMAP v1.1 (41) was used to pinpoint putatively causal SNPs. These were defined as individual SNPs with posterior probability ($PP_{FINEMAP}$) higher than 95%. In the absence of such SNPs, a list of credible SNPs was generated, which included those with cumulative $PP_{FINEMAP}$ = 95%. Sets of credible SNPs were annotated to function using the Ensembl Variant Effect Predictor (VEP) tool (42). To attempt a more accurate identification of putatively causal genes in these loci, we also analysed gene expression from multiple tissues, including liver, using data from GTEx v7 (43) (**Supplementary Methods**). Information about hepatic promoters, enhancers and topologically-associated domains (TADs) was retrieved from (44) and added to the SNP-based annotations.

Estimating the effect of individual SNPs on plasma concentration.

The modelling we undertook to derive our primary GWAS phenotype produces genetic effect sizes in LMM-LOCO that are related to the residuals used as our primary outcome rather than to the raw assay data, which would be easily interpretable. To explicitly estimate genetic effects on the scale of our clozapine metabolite plasma concentrations, we extracted the minor allele counts ("allelic dosages") for genome-wide significant SNPs of each GWAS for each individual. Mixed regression models were fitted for each outcome variable including as covariates clozapine dose, time since dose, age, sex, allelic dosage and the first 20 principal components (PCs) derived from the genotype data using PC-AiR (45). A random effect covariate was used to capture individual-level variance in this model. The effect of allelic dosage on plasma concentration was then estimated within the same mixed linear regression framework used to generate the GWAS phenotypes, which produces a more meaningful result from the clinical point of view. Model fitting statistics (e.g. variance explained by fixed and random effects, proportion of variance explained by single SNPs) were determined as described in **Supplementary Methods**.

Locating shared associations with other metabolic traits

We used GWAS-pw v0.21 (46) to identify genetic markers associated with clozapine and norclozapine concentrations that were also associated with the concentrations of other metabolites in the KORA/TWINSUK study (47) in a genome-wide context. We focused on the 285 metabolites and xenobiotics confidently identified in that study, disregarding unknown compounds and metabolite ratios. Using the summary statistics from a GWAS of each metabolite, we generated co-localisation posterior probabilities (PP_{COLOC}) with our clozapine and norclozapine summary statistics. Probabilities were calculated for all individual SNPs with complete data, in order to identify shared effects inside and outside genome-wide significant loci, thus allowing us to pinpoint more robust shared effects than would emerge from examining a limited number of loci.

Analysis of human metabolic pathways

To study the genetic component of clozapine metabolism in the context of the human metabolic network, we retrieved the most recent metabolome reconstruction, RECON 2.2, capturing 5,324 metabolites and 1,675 genes (48). We grouped genes into subsystems (e.g. "extracellular transport", "steroid metabolism"), resulting in 79 gene-sets. One additional set was created from 203 genes analysed in a recent drug metabolism study (49), representing known pharmacokinetic-relevant enzymes and receptors. Gene-set enrichment analysis was performed with MAGMA v1.06 (50), using the "multi" method to calculate gene-wide p-values from GWAS summary statistics. Within each analysis, gene-set p-values were corrected using the family-wise error rate (FWER) with 100,000 permutations.

Results

Genome-wide significant SNPs associated with clozapine plasma concentrations

The GWAS of clozapine levels identified a single genome-wide significant association at rs2472297, an intergenic variant between *CYP1A1* and *CYP1A2* (**Figure 1A; Table 1; Supplementary Table 1; Supplementary Figure 2**). Analysis of GTEx hepatocyte expression data did not relate this signal to any particular gene (**Supplementary Methods**), though rs2472297 has been previously associated with CYP1A2 activity on the basis of its effect on caffeine metabolite concentrations (51). In the mixed-model analysis, the minor allele of this variant was shown to be associated with reduced clozapine plasma concentrations, with a proportion of variance

explained (PVE) of 1.47% (**Table 2, Figure 2, Supplementary Methods**). Model-fitting statistics for the complete groups of fixed and random effects are listed in **Supplementary Table 2**.

The GWAS of norclozapine levels identified two genome-wide significant loci (**Figure 1B; Table 1; Supplementary Table 1; Supplementary Figure 3**). The first was indexed by rs72846859, an intergenic variant upstream of *UGT2B10*. FINEMAP revealed a complex association signal in this region, with 171 credible SNPs (**Supplementary Table 3**) including a missense variant (Asp/Tyr) in *UGT2B10*, rs61750900. LD between the index and missense variants was high (r^2= 0.964), and given its higher prior probability of causality (**Supplementary Methods**) we incorporated rs61750900 into the mixed regression model of norclozapine plasma levels. For the second genome-wide significant locus, the index SNP was rs2011425, a missense variant (Leu/Val) in *UGT1A4*, which also obtained the highest FINEMAP probability out of 47 credible SNPs (**Supplementary Table 3**). The minor alleles of both SNPs were associated with lower norclozapine plasma levels, with a PVE=2.32% for rs61750900 and PVE=1.15% for rs2011425 (**Table 2**).

The GWAS of clozapine/norclozapine metabolic ratio identified three independent genome-wide significant associations at two distinct loci (**Figure 1C; Table 1; Supplementary Table 1; Supplementary Figure 4**). Two of these LD-independent SNPs (rs10023464, rs7668556) tagged a locus on chromosome 4 that includes seven genes of the *UGT2* family. The remaining SNP, rs12767583, is an intronic variant in *CYP2C19*. FINEMAP showed that both loci harbour complex association signals, returning 65 and 102 credible SNPs, respectively (**Supplementary Table 4**). At each locus, the set of credible SNPs included a missense variant in high LD (r^2>0.9) with the top FINEMAP SNP, one of which (rs61750900) was also genome-wide significant in the norclozapine analysis. Both missense variants, rs61750900 (PVE=9.48%) and rs1126545 (*CYP2C18* Thr/Met; PVE=1.85%), were incorporated into a log-normal model, where their minor allele dosage was shown to increase the clozapine/norclozapine ratio (**Table 2**).

Secondary GWAS analyses controlling for smoking and BMI PRS gave very similar results to those reported here, with no gain or loss of genome-wide significant signals (**Supplementary Methods**). Also, confirming observations from a previous study conducted using multiple assays (52), we find that our approach of using mixed model residuals as a GWAS phenotype gives results similar to the use of summary statistics (averages or maximum values), but with tighter standard errors resulting in improved significance for individual loci (**Supplementary Figure 6**).

Co-localisation analysis of metabolite levels

We employed a co-localisation procedure to test whether SNPs implicated in clozapine levels might also impact the plasma concentrations of other compounds, as this can provide insight into the causal mechanisms behind these signals and reveal metabolic convergences and potential clinically important interactions. Analysis of the clozapine levels GWAS showed that the association at the *CYP1A2* locus, indexed by rs2472297, was also observed (PP_{COLOC} > 94%) in GWAS of five xenobiotic metabolites: caffeine, theophylline, 7-methylxanthine, paraxanthine, and Leu-Pro cyclopeptide. All of these are putative biomarkers of coffee consumption (53): the first four are implicated in caffeine metabolism, while Leu-Pro cyclopeptide is a component of roasted coffee. On the basis of these results, we carried out a polygenic score analysis to obtain a surrogate metric of daily coffee intake (**Supplementary Methods**), which we found to be significantly associated with all of our phenotypes (**Supplementary Table 5**).

Analysing the norclozapine levels GWAS, the xenobiotic metabolite pelargonate co-localised at the main *UGT2B10* locus (PP_{COLOC} = 96.97%), while caffeine and theophylline co-localised at the *CYP1A1/CYP1A2* locus (PP_{COLOC} > 98%), which for norclozapine is indexed by rs2472297 ($P=3.52x10^{-5}$). Although pelargonate is a component of some commercial coffee varieties, it is not part of the caffeine metabolic pathway, but of the wider system of fatty acid metabolism, as is UGT2B10 (54). Interestingly, pelargonate shares structural similarities with the antiepileptic valproate, and it has recently been shown that valproate co-administration with clozapine reduces norclozapine plasma levels (55).

Genome-wide enrichment of metabolic pathways

After correction for multiple testing, 5 gene-sets were significant in our analysis of RECON biochemical pathways (**Table 3, Supplementary Table 6**). Vitamin A (retinol) metabolism was the top enriched pathway for both clozapine and norclozapine, while linoleate metabolism was the second norclozapine pathway and the top pathway for the clozapine/norclozapine ratio. In clozapine and norclozapine, we also observed significant FWER-corrected enrichment for the set of 203 drug metabolising enzymes; repeating the enrichment analysis using this gene-set as a covariate removed all other gene-set signals from the clozapine GWAS, while vitamin A and linoleate (a fatty acid) remained significant in in the gene-set analyses of norclozapine and the metabolic ratio (**Supplementary Table 7**).

Discussion

We have carried out the first GWAS of clozapine metabolite plasma concentrations in 2,989 European individuals, the majority of whom had been assayed at multiple time points. Using statistical modelling to take advantage of all the available data, and a linear mixed-model GWAS approach, we provide the first robust evidence that alleles of specific *CYP* and *UGT* genes contribute to clozapine pharmacokinetics. This represents an advance from previous inconclusive studies (28), mostly based on candidate marker surveys, and clarifies the relevance of common genetic variation in the proteins implicated in the clozapine metabolic route, which has been a matter of extensive debate. More specifically, our results support the hypothesis that the genetic architecture of clozapine metabolism might be driven by a few variants of large effect (**Supplementary Figure 7**), in line with other well-studied metabolic traits (56).

The CYP1A1/CYP1A2 SNP rs2472297 associated with clozapine plasma concentrations lies in an intergenic region rich in binding sites for the Aryl Hydrocarbon Receptor (AHR) protein, sites which are collectively known as "xenobiotic response elements" (57). AHR binding is known to induce the expression of CYP enzymes in hepatocytes in response to the detection of many compounds, and thus variation in the regulatory function of AHR provides a strong candidate mechanism underpinning this association. While a causal association cannot be made solely on these grounds, disruption of normal AHR binding has also been suggested as explaining the association between variants at this locus and caffeine plasma levels, which may also influence coffee consumption (51). Previous studies of clozapine levels and candidate polymorphisms at this locus have focussed on common alleles within *CYP1A2* (27, 28), none of which has been shown to influence its expression (58). We also note we find no support for other candidate genes from the literature, including the *ABCB1* variant rs1045642 (P=0.84) which was previously reported as associated with clozapine plasma concentrations in smaller (N<100) samples (28). Both of these examples demonstrate the limitations of candidate SNP approaches that have been common in psychiatric pharmacogenomics to date and support genome-wide analysis to capture both coding and non-coding functional elements.

The results of our regression modelling show that the genetic modifiers of clozapine levels are comparable in impact to other known clinical and demographic variables, with their effect sizes being of the same magnitude as sex (**Table 1**, **Table 2)**. An example to illustrate these effects and place them in clinical context is the observation that carrying one minor allele of rs2472297 at *CYP1A1/CYP1A2* is associated with a reduction in clozapine plasma concentrations roughly equivalent to a decrease in clozapine by 50 mg/day, and homozygosity for the minor allele is

equivalent to a reduction by 100mg (**Figure 2**). Similar effects were found for the missense SNPs associated with norclozapine levels (**Supplementary Figure 8**). The impacts on clozapine metabolite concentrations captured by these SNPs warrants their further study within the context of personalised drug therapy, given their potential clinically significant impact on dosing.

In following up the results of the GWAS we sought genomic regions associated with clozapine metabolism that have also been identified as influencing metabolism of other compounds. We identified a strong relationship between the genetics of clozapine and caffeine metabolism, a finding with potential clinical relevance. A link between clozapine and caffeine metabolism was first proposed on the basis that the results of caffeine clearance tests, used as an index of CYP1A2 activity, correlate with clozapine clearance (59). Whilst there have not been large-scale studies in clinical settings, the available evidence suggests that caffeine interacts competitively with clozapine, causing heavy coffee drinkers to have higher baseline clozapine plasma levels (60). Among factors that may have obscured this finding in previous research are the proven correlation between smoking and coffee consumption (61), and the observation that even decaffeinated coffee might lower the activity of some hepatic enzymes (62). In this regard, our analysis of metabolic-genetic association data showed commonalities, which hint to potential interactions, with several compounds related to coffee and caffeine. Remarkably, loci outside of the widely-studied CYP1A2 region seem to jointly impact both coffee consumption habits and plasma concentration of clozapine metabolites, as we have shown using a polygenic score approach (**Supplementary Methods**). However, given that we did not have access to coffee or caffeine consumption data, we could not assess the degree to which caffeine may be mediating or moderating the genetic associations with clozapine metabolite levels. Nonetheless our results add to the existing evidence of the potential clinical importance of the interaction between the metabolic pathways for clozapine and caffeine.

Our data can also be interpreted in the light of the proposed mechanistic link between smoking tobacco and clozapine metabolism, which is thought to result from induction by tobacco of CYP1A2 activity, which in turn increases the first-pass metabolism of clozapine (63). Current guidelines state that patients on clozapine need to be more closely monitored if they stop smoking, as their plasma levels can suddenly rise as the CYP1A2 induction fades. This effect, also seen with other medications, has been attributed to the effect of polycyclic hydrocarbons present in tobacco smoke, rather than a direct action of nicotine (64). As such, non-smoke alternatives to tobacco, such as nicotine patches or e-cigarettes, are generally considered not likely to interact

with clozapine treatment. However, we have shown that genetic variants in UGT enzymes, which are responsible for nicotine glucuronidation, also have a role in the clozapine metabolism. Specifically, we have highlighted a missense polymorphism in UGT2B10, previously shown to result in impaired enzymatic function (65), as a credible causal variant for influencing norclozapine plasma levels. This enzyme has also been shown to be a substrate of several antipsychotic drugs with similar structural properties to clozapine (66). Given that nicotine is a specific high-affinity inhibitor of UGT2B10, our results support the possibility of nicotine-clozapine interactions in the glucuronidation excretion pathway (26, 66), which should be investigated in more detail.

One of the limitations of this study is that our regression models do not explain as much variance in plasma concentrations as previous studies (16, 17). However, we note that these included the clozapine/norclozapine metabolic ratio as a covariate of clozapine plasma concentrations. Given our data, and considering all fixed effect covariates (**Supplementary Table 2**), this addition would have increased the variance explained by our mixed model from 19.28% to 32.34%, but at the cost of adding collinearity and hindering its interpretability. In any case, these models likely represent a lower bound of variance explained for clozapine plasma concentrations, given we lacked individual measures of some known predictors of clozapine metabolism, including smoking habit and weight. We have attempted to address this limitation in the discovery GWAS by using a novel application of PRS as genetically informative proxies of these measures (**Supplementary Methods**). While this did not impact the results, we acknowledge these are just markers of the exposures and do not capture their full effects. A further limitation is the lack of detailed individual level data on concomitant medications that could interact with clozapine. Co-prescription of such medications (e.g. carbamazepine and fluvoxamine) has been shown to be rare given their potential for clinically important interactions (67), and hence it does not seem feasible that such co-prescription could be an important source of bias in our findings. Furthermore, given that the absence of detailed individual-level exposure data is known to obscure the detection of genetic influences in metabolic enzymes (68), our finding of detectable GWAS signals is reassuring. A final limitation is the potential that those who have had their clozapine levels taken might be an unrepresentative sample of all those taking clozapine, which would constitute a form of selection bias. In examining this issue we did not detect differences between those with or without clozapine TDM assays in the distribution of age, gender and several PRS (schizophrenia, IQ, BMI, smoking). Nonetheless we cannot rule out other selection effects, and thus our findings should be interpreted as relevant to the population in which clozapine TDM levels are monitored.

In summary, our analysis has allowed us to dissect the clozapine metabolic pathway using genetic and pharmacokinetic data. We have also demonstrated commonalities with the metabolism of other biological compounds, in particular nicotine and caffeine, which highlight relevant facets of metabolism and indicate potential interactions of clinical importance. Furthermore, our findings indicate avenues for next-stage clinical studies to determine the utility of pharmacogenomic testing as a complement to clozapine monitoring procedures, with the potential to impact clinical care through improved titration, dosing, and minimising of ADRs.

Table 1: Association statistics of the index SNPs for each phenotype and LD-independent locus. Results for fine-mapped missense variants in high LD with each index SNP are also shown.

Phenotype	Locus	SNP	Allele	Annotation	GWAS P-value
Clozapine	chr15:74817689-75404506	rs2472297	T	Intergenic	4.35×10^{-10}
Norclozapine	chr4:69542100-70312793	rs11725502	T	Intergenic	5.47×10^{-15}
Norclozapine	chr4:69542100-70312793	rs61750900	T	Missense	8.91×10^{-15}
Norclozapine	chr2:234611523-234676118	rs2011425	G	Missense	8.37×10^{-9}
Ratio	chr4:69542100-70387482	rs10023464	T	Intergenic	8.72×10^{-66}
Ratio	chr4:69542100-70387482	rs61750900	T	Missense	1.69×10^{-64}
Ratio	chr10:96098093-96974830	rs12767583	T	Intronic	4.64×10^{-14}
Ratio	chr10:96098093-96974830	rs1126545	T	Missense	1.02×10^{-13}

Table 2: Effect sizes of genetic, demographic and clinical covariates as estimated with linear mixed regression modelling. All models also included age^2 and 20 genotype principal components as fixed effects (omitted)

Phenotype	Covariate	Beta	Standard error	P-value
Clozapine	rs2472297 (T)	-0.089	0.013	2.40×10^{-11}
Clozapine	Clozapine daily dose (mg)	0.002	4.03×10^{-5}	$<1 \times 10^{-300}$
Clozapine	Time since last dose (hours)	-0.009	0.002	8.11×10^{-6}
Clozapine	Patient age (years)	0.004	0.004	0.384
Clozapine	Patient gender (reference=male)	-0.147	0.019	1.31×10^{-14}
Norclozapine	rs61750900 (T)	-0.149	0.018	3.17×10^{-17}
Norclozapine	rs2011425 (G)	-0.112	0.019	3.34×10^{-9}
Norclozapine	Clozapine daily dose (mg)	0.002	3.67×10^{-5}	$<1 \times 10^{-300}$
Norclozapine	Time since last dose (hours)	6.81×10^{-4}	0.002	0.701
Norclozapine	Patient age (years)	-0.003	0.004	0.444
Norclozapine	Patient gender (reference=male)	-0.120	0.017	3.61×10^{-12}
Ratio	rs61750900 (T)	0.212	0.012	5.01×10^{-70}
Ratio	rs1126545 (T)	0.078	0.010	5.96×10^{-14}
Ratio	Clozapine daily dose (mg)	-1.49×10^{-4}	2.40×10^{-5}	5.22×10^{-10}
Ratio	Time since last dose (hours)	-0.014	0.001	1.03×10^{-33}
Ratio	Patient age (years)	0.007	0.003	0.006
Ratio	Patient gender (reference=male)	-0.016	0.012	0.164

Table 3: Gene sets surviving FWER correction (P_{FWER} > 0.05) from the MAGMA gene set analysis of the RECON metabolic pathways.

Phenotype	Gene set	N_{GENES}	β	s.e.	P_{MAGMA}	P_{FWER}
Clozapine	Retinol metabolism	33	0.635	0.18	$2.07x10^{-4}$	0.016
Norclozapine	Retinol metabolism	33	1.08	0.0448	$3.95x10^{-10}$	$3.12x10^{-8}$
Norclozapine	Linoleate metabolism	16	1.5	0.0433	$2.30x10^{-7}$	$6.00x10^{-5}$
Norclozapine	Arachidonate metabolism	22	0.552	0.0187	$1.17x10^{-4}$	0.009
Norclozapine	Steroid metabolism	41	0.528	0.0244	$1.53x10^{-4}$	0.012
Ratio	Linoleate metabolism	16	1.54	0.0445	$5.95x10^{-7}$	$1.30x10^{-4}$
Ratio	Chondroitin degradation	10	0.753	0.0172	$2.07x10^{-4}$	0.039

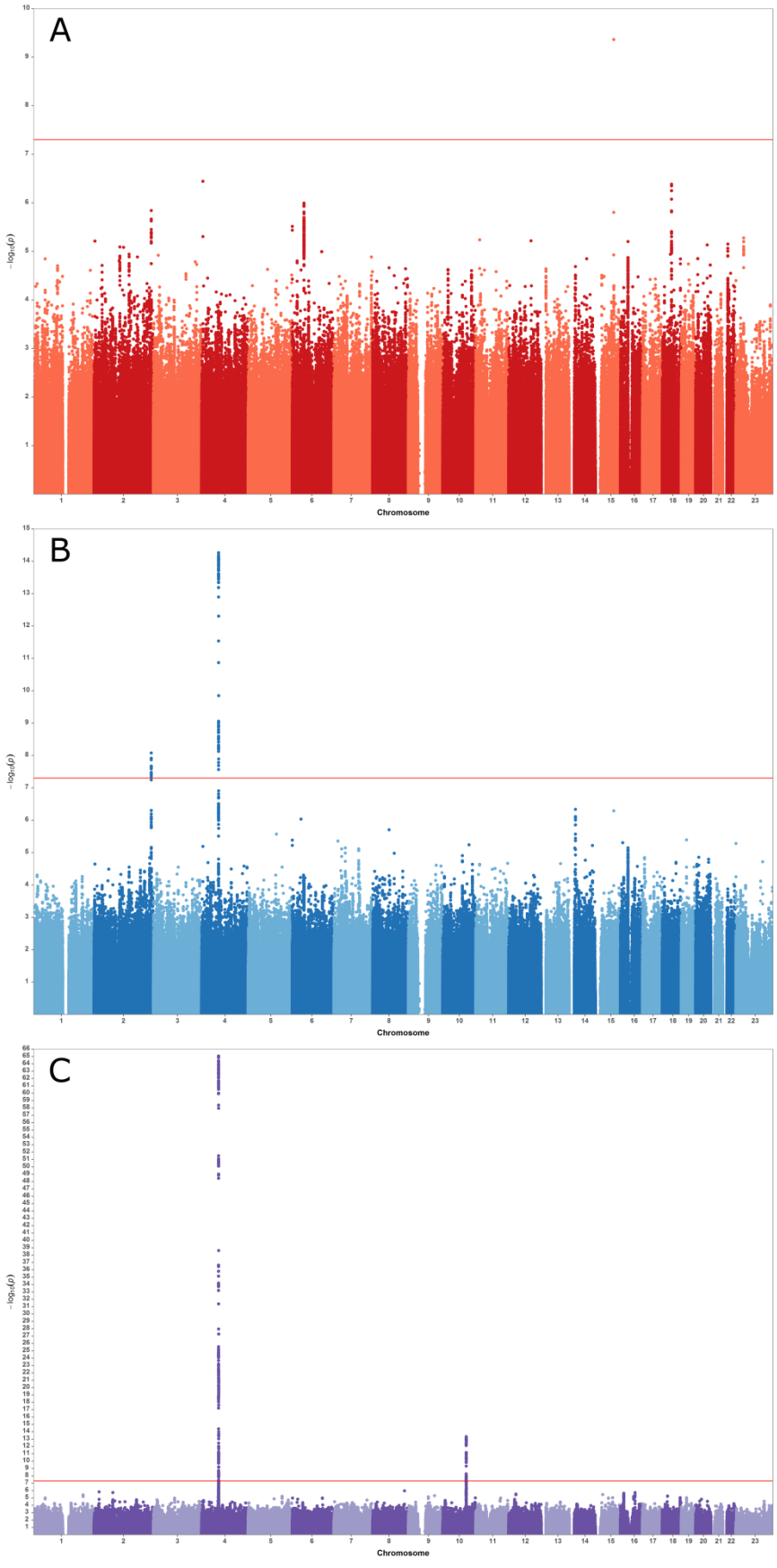

Figure 1: Manhattan plots of the clozapine (A), norclozapine (B) and metabolic ratio (C) GWAS.

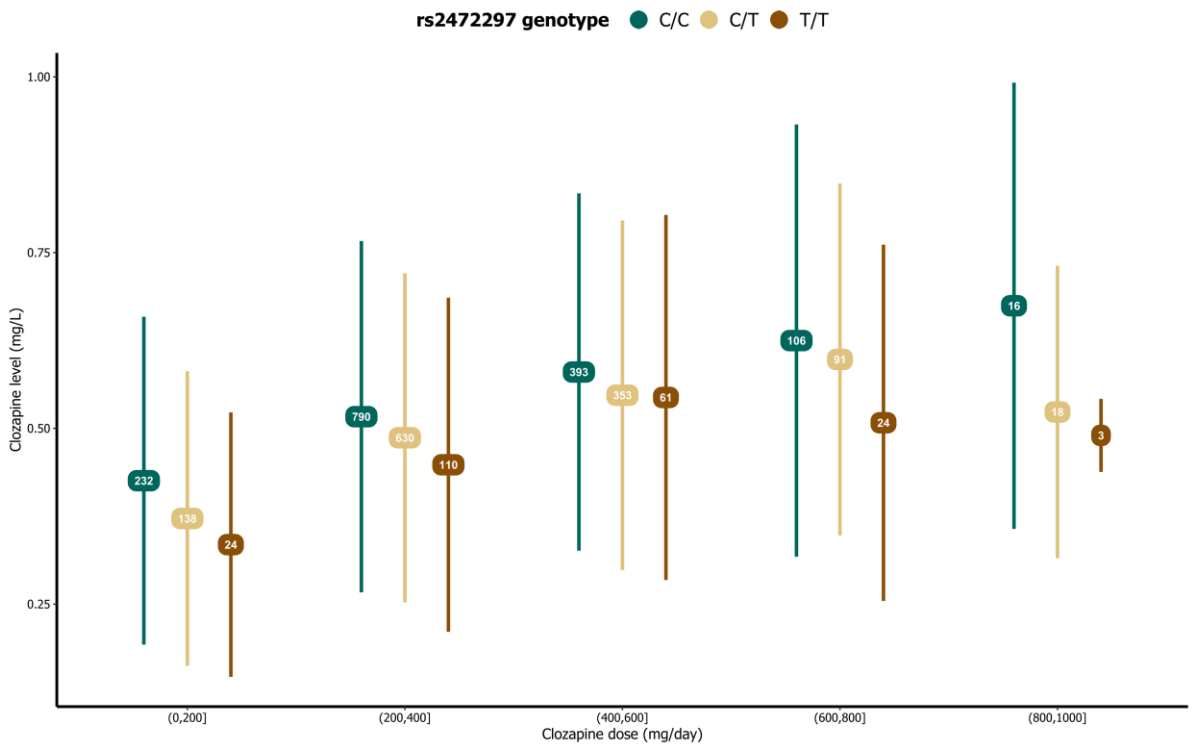

Figure 2: Effect of the rs2472297 genotype on clozapine plasma levels, at different daily clozapine doses. For this analysis, only the last time point of each CLOZUK2 individual was used. For each interval of daily clozapine dose, average plasma concentrations and standard deviations are shown. Values inside the central point represent the number of individuals within each genotype/interval category.

References

1. McGrath J, Saha S, Chant D, et al.: Schizophrenia: A Concise Overview of Incidence, Prevalence, and Mortality. **Epidemiol Rev** 2008;30:67-76.

2. Owen MJ, Sawa A, Mortensen PB: Schizophrenia. **Lancet** 2016;388:86–97.

3. Howes OD, McCutcheon R, Agid O, et al.: Treatment-Resistant Schizophrenia: Treatment Response and Resistance in Psychosis (TRRIP) Working Group Consensus Guidelines on Diagnosis and Terminology. **Am J Psychiatry** 2017;174:216-229.

4. Essali A, Al-Haj Haasan N, Li C, et al.: Clozapine versus typical neurolepticmedication for schizophrenia. **Cochrane Database Syst Rev** 2009;CD000059.

5. Kelly DL, Wehring HJ, Vyas G: Current status of clozapine in the United States. **Shanghai Arch Psychiatry** 2012;24:110-113.

6. Latimer E, Wynant W, Clark R, et al.: Underprescribing of Clozapine and Unexplained Variation in Use across Hospitals and Regions in the Canadian Province of Québec. **Clin Schizophr Relat Psychoses** 2013;7:33-41.

7. Atkin K, Kendall F, Gould D, et al.: Neutropenia and agranulocytosis in patients receiving clozapine in the UK and Ireland. **Br J Psychiatry** 1996;169:483-488.

8. Nair B, MacCabe JH: Making clozapine safer: current perspectives on improving its tolerability. **Future Neurol** 2014;9:313-322.

9. Best-Shaw L, Gudbrandsen M, Nagar J, et al.: Psychiatrists' Perspectives on Antipsychotic Dose and the Role of Plasma Concentration Therapeutic Drug Monitoring. **Ther Drug Monit** 2014;36:486-493.

10. Legge SE, Hamshere M, Hayes RD, et al.: Reasons for discontinuing clozapine. **Schizophr Res** 2016;174:113-119.

11. Flanagan RJ: A practical approach to clozapine therapeutic drug monitoring. **CMHP Bulletin** 2010;2:4-5.

12. Spina E, Hiemke C, de Leon J: Assessing drug-drug interactions through therapeutic drug monitoring when administering oral second-generation antipsychotics. **Expert Opin Drug Metab Toxicol** 2016;12:407-422.

13. Hiemke C, Bergemann N, Clement HW, et al.: Consensus Guidelines for Therapeutic Drug Monitoring in Neuropsychopharmacology: Update 2017. **Pharmacopsychiatry** 2018;51:9-62.

14. Simon V, Ruud van Winkel M, Marc De Hert M: Are Weight Gain and Metabolic Side Effects of Atypical Antipsychotics Dose Dependent? A Literature Review. **J Clin Psychiatry** 2009;70:1041-1050.

15. Ramos Perdigués S, Sauras Quecuti R, Mané A, et al.: An observational study of clozapine induced sedation and its pharmacological management. **Eur Neuropsychopharmacol** 2016;26:156-161.

16. Rostami-Hodjegan A, Amin AM, Spencer EP, et al.: Influence of dose, cigarette smoking, age, sex, and metabolic activity on plasma clozapine concentrations: a predictive model and nomograms to

aid clozapine dose adjustment and to assess compliance in individual patients. **J Clin Psychopharmacol** 2004;24:70-78.

17. Couchman L, Bowskill SVJ, Handley S, et al.: Plasma clozapine and norclozapine in relation to prescribed dose and other factors in patients aged <18 years: data from a therapeutic drug monitoring service, 1994–2010. **Early Interv Psychiatry** 2013;7:122-130.

18. Russell Teagarden J, Stanek EJ: On Pharmacogenomics in Pharmacy Benefit Management. **Pharmacotherapy** 2012;32:103-111.

19. Lauschke VM, Milani L, Ingelman-Sundberg M: Pharmacogenomic Biomarkers for Improved Drug Therapy—Recent Progress and Future Developments. **The AAPS Journal** 2017;20:4.

20. de Leon J: Have we successfully implemented CYP2D6 genotyping in psychiatry? **Expert Opin Drug Metab Toxicol** 2017;13:1201-1203.

21. Jann MW, Grimsley SR, Gray EC, et al.: Pharmacokinetics and Pharmacodynamics of Clozapine. **Clin Pharmacokinet** 1993;24:161-176.

22. Eiermann B, Engel G, Johansson I, et al.: The involvement of CYP1A2 and CYP3A4 in the metabolism of clozapine. **Br J Clin Pharmacol** 1997;44:439-446.

23. Lameh J, Burstein ES, Taylor E, et al.: Pharmacology of N-desmethylclozapine. **Pharmacol Ther** 2007;115:223-231.

24. Breyer-Pfaff U, Wachsmuth H: Tertiary N-glucuronides of clozapine and its metabolite desmethylclozapine in patient urine. **Drug Metab Dispos** 2001;29:1343-1348.

25. Zanger UM, Schwab M: Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. **Pharmacol Ther** 2013;138:103-141.

26. Rowland A, Miners JO, Mackenzie PI: The UDP-glucuronosyltransferases: Their role in drug metabolism and detoxification. **Int J Biochem Cell Biol** 2013;45:1121-1132.

27. Bersani FS, Capra E, Minichino A, et al.: Factors affecting interindividual differences in clozapine response: a review and case report. **Hum Psychopharmacol** 2011;26:177-187.

28. Krivoy A, Gaughran F, Weizman A, et al.: Gene polymorphisms potentially related to the pharmacokinetics of clozapine: a systematic review. **Int Clin Psychopharmacol** 2016;31:179-184.

29. Rietveld CA, Conley D, Eriksson N, et al.: Replicability and Robustness of Genome-Wide-Association Studies for Behavioral Traits. **Psychological Science** 2014;25:1975-1986.

30. Pardiñas AF, Holmans P, Pocklington AJ, et al.: Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. **Nat Genet** 2018;50:381–389.

31. Chang CC, Chow CC, Tellier L, et al.: Second-generation PLINK: rising to the challenge of larger and richer datasets. **Gigascience** 2015;4.

32. Anderson CA, Pettersson FH, Clarke GM, et al.: Data quality control in genetic case-control association studies. **Nat Protoc** 2010;5:1564–1573.

33. McCarthy S, Das S, Kretzschmar W, et al.: A reference panel of 64,976 haplotypes for genotype imputation. **Nat Genet** 2016;48:1279-1283.

34. Das S, Forer L, Schonherr S, et al.: Next-generation genotype imputation service and methods. **Nat Genet** 2016;48:1284-1287.

35. Advanced Research Computing @ Cardiff (ARCCA). **Introduction to RAVEN**. Retrieved from http://www.cardiff.ac.uk/arcca/services/equipment/ravenintroduction.html

36. van Leeuwen EM, Kanterakis A, Deelen P, et al.: Population-specific genotype imputations using minimac or IMPUTE2. **Nat Protocols** 2015;10:1285-1296.

37. Huang J, Howie B, McCarthy S, et al.: Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. **Nat Commun** 2015;6:8111.

38. Legge SE, Pardiñas AF, Helthuis M, et al.: A genome-wide association study in individuals of African ancestry reveals the importance of the Duffy-null genotype in the assessment of clozapine-related neutropenia. **Mol Psychiatry** 2018;[in press].

39. Grundmann M, Kacirova I, Urinovska R: Therapeutic drug monitoring of atypical antipsychotic drugs. **Acta Pharm** 2014;64:387-401.

40. Yang J, Lee SH, Goddard ME, et al.: GCTA: A Tool for Genome-wide Complex Trait Analysis. **Am J Hum Genet** 2011;88:76-82.

41. Benner C, Spencer CCA, Havulinna AS, et al.: FINEMAP: efficient variable selection using summary data from genome-wide association studies. **Bioinformatics** 2016;32:1493-1501.

42. McLaren W, Pritchard B, Rios D, et al.: Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. **Bioinformatics** 2010;26:2069-2070.

43. Carithers LJ, Ardlie K, Barcus M, et al.: A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. **Biopreserv Biobank** 2015;13:311-319.

44. Schmitt AD, Hu M, Jung I, et al.: A Compendium of Chromatin Contact Maps Reveal Spatially Active Regions in the Human Genome. **Cell reports** 2016;17:2042-2059.

45. Conomos MP, Miller MB, Thornton TA: Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. **Genet Epidemiol** 2015;39:276-293.

46. Pickrell JK, Berisa T, Liu JZ, et al.: Detection and interpretation of shared genetic influences on 42 human traits. **Nat Genet** 2016;48:709-717.

47. Shin S-Y, Fauman EB, Petersen A-K, et al.: An atlas of genetic influences on human blood metabolites. **Nat Genet** 2014;46:543-550.

48. Swainston N, Smallbone K, Hefzi H, et al.: Recon 2.2: from reconstruction to model of human metabolism. **Metabolomics** 2016;12:109.

49. Schärfe CPI, Tremmel R, Schwab M, et al.: Genetic variation in human drug-related genes. **Genome Med** 2017;9:117.

50. de Leeuw CA, Mooij JM, Heskes T, et al.: MAGMA: Generalized gene-set analysis of GWAS data. **PLoS Comput Biol** 2015;11:e1004219.

51. Cornelis MC, Kacprowski T, Menni C, et al.: Genome-wide association study of caffeine metabolites provides new insights to caffeine metabolism and dietary caffeine-consumption behavior. **Hum Mol Genet** 2016;25:5472-5482.

52. Hoffmann TJ, Theusch E, Haldar T, et al.: A large electronic-health-record-based genome-wide study of serum lipids. **Nat Genet** 2018;50:401-413.

53. Guertin KA, Loftfield E, Boca SM, et al.: Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. **Am J Clin Nutr** 2015;101:1000-1011.

54. Turgeon D, Chouinard S, Bélanger P, et al.: Glucuronidation of arachidonic and linoleic acid metabolites by human UDP-glucuronosyltransferases. **J Lipid Res** 2003;44:1182-1191.

55. Hommers L, Scharl M, Hefner G, et al.: Comedication of Valproic Acid Is Associated With Increased Metabolism of Clozapine. **J Clin Psychopharmacol** 2018;38:188-192.

56. Timpson NJ, Greenwood CMT, Soranzo N, et al.: Genetic architecture: the shape of the genetic contribution to human traits and disease. **Nature Rev Genet** 2017;19:110.

57. Corchero J, Pimprale S, Kimura S, et al.: Organization of the CYP1A cluster on human chromosome 15: implications for gene regulation. **Pharmacogenet Genomics** 2001;11:1-6.

58. Ingelman-Sundberg M, Sim SC, Gomez A, et al.: Influence of cytochrome P450 polymorphisms on drug therapies: Pharmacogenetic, pharmacoepigenetic and clinical aspects. **Pharmacol Ther** 2007;116:496-526.

59. Bertilsson L, Carrillo JA, Dahl ML, et al.: Clozapine disposition covaries with CYP1A2 activity determined by a caffeine test. **Br J Clin Pharmacol** 1994;38:471-473.

60. Raaska K, Raitasuo V, Laitila J, et al.: Effect of Caffeine-Containing versus Decaffeinated Coffee on Serum Clozapine Concentrations in Hospitalised Patients. **Basic Clin Pharmacol Toxicol** 2004;94:13-18.

61. Treur JL, Taylor AE, Ware JJ, et al.: Associations between smoking and caffeine consumption in two European cohorts. **Addiction** 2016;111:1059-1068.

62. Xiao Q, Sinha R, Graubard BI, et al.: Inverse associations of total and decaffeinated coffee with liver enzyme levels in National Health and Nutrition Examination Survey 1999-2010. **Hepatology** 2014;60:2091-2098.

63. Lowe EJ, Ackman ML: Impact of tobacco smoking cessation on stable clozapine or olanzapine treatment. **Ann Pharmacother** 2010;44:727-732.

64. Smith TL, Mican LM: What to do when your patient who takes clozapine enters a smoke-free facility. **Current Psychiatry** 2014;13:55.

65. Chen G, Blevins-Primeau AS, Dellinger RW, et al.: Glucuronidation of Nicotine and Cotinine by UGT2B10: Loss of Function by the UGT2B10 Codon 67 (Asp/Tyr) Polymorphism. **Cancer Res** 2007;67:9024-9029.

66. Pattanawongsa A, Nair PC, Rowland A, et al.: Human UDP-Glucuronosyltransferase (UGT) 2B10: Validation of Cotinine as a Selective Probe Substrate, Inhibition by UGT Enzyme-Selective Inhibitors and Antidepressant and Antipsychotic Drugs, and Structural Determinants of Enzyme Inhibition. **Drug Metab Dispos** 2016;44:378-388.

67. Heald A, Livingston M, Yung A, et al.: Prescribing in schizophrenia and psychosis: Increasing polypharmacy over time. **Hum Psychopharmacol** 2017;32:e2579.

68. Matthaei J, Tzvetkov M, Strube J, et al.: Heritability of Caffeine Metabolism: Environmental Effects Masking Genetic Effects on CYP1A2 Activity but Not on NAT2. **Clin Pharmacol Ther** 2016;100:606-616.