Data Supplement for Pardiñas et al., Pharmacogenomic Variants and Drug Interactions Identified Through the Genetic Analysis of Clozapine Metabolism. Am J Psychiatry (doi: 10.1176/appi.ajp.2019.18050589)

SUPPLEMENTARY METHODS

Supplementary Tables S1–S7 are in a separate Excel file.

Analysing plasma samples of TRS patients to obtain clozapine metabolite data

Plasma from the CLOZUK2 individuals was initially separated from whole blood by centrifugation, and the top plasma layer (minimum 200µL of plasma) was removed for analysis. Samples were prepared in 96-well batches with duplicate calibrators and multiple quality control materials, including samples of known concentration. Chromatography was performed on a Waters CSH Phenyl-Hexyl, 50mm x 3mm, reverse phase HPLC column. After mass spectrometry, total abundance of each of the product ions was monitored and plotted to produce a chromatogram, and the area of each peak was divided by the area of the corresponding internal standard. This ratio was used to determine the concentration when compared to that of the calibrators. Calibration curves had a range of 0.05mg/L to 3mg/L for both clozapine and norclozapine.

Combining data from multiple metabolite assays into a single phenotype per individual

We first assessed the distribution of clozapine and norclozapine concentrations, using the "*fitdistrplus"* v1.09 R package (1). This is necessary as pharmacokinetic variables often do not conform to normal distributions (2), which has been shown specifically for clozapine (3). We examined the set of distributions highlighted by Lindsey et al. 2001 (2); which included the normal, log-normal, gamma, Weibull, log-Laplace and log-Cauchy. For both clozapine and norclozapine plasma concentrations, the best fitting distribution was the gamma, chosen as that with the minimal Anderson-Darling distance (4). Testing the above distributions, as well as the beta prime (ratio of two gamma variables) identified the log-normal distribution as best fitting for the metabolic ratio.

A generalised linear model was fitted on the assay data for each outcome variable (clozapine, norclozapine and the metabolic ratio) using the "*gamlss"* v5.02 R package (5). The appropriate distribution as determined above was considered. The fixed effect covariates incorporated were dose of clozapine, time between dose and assay, age (at the time of each assay) and $(age)^2$. A random effect was added to model the distribution of the outcome in each individual, controlled for the covariates. For each outcome variable, the coefficients of this random effect for each individual were extracted, and considered as the GWAS phenotype.

Assessing the variance explained by a mixed linear model

As has been extensively discussed before (6), estimating a coefficient of determination (R^2) that captures the proportion of variance explained by a mixed model is a non-trivial issue, since the presence of random

effect factors complicates most variance decomposition procedures. Our reported values are based in the approach of Nakagawa et al. (7, 8), since its formulae are valid for both log-normal and gamma-distributed outcome variables, such as those described in the main text. Specifically, reported R^2 values correspond to the $R^2_{GLMM(m)}$ approach, which captures the proportion of variance explained by the fixed effect factors in a mixed model design, also termed as "marginal effects". Reported variances explained by the random effect factors correspond to the $ICC_{adjusted}$ approach, which in our case estimates the proportion of the total variance explained by grouping the repeated measures in individuals. Note that this is equivalent to the statistic termed "intra-class correlation coefficient" or "repeatability R^2" in other publications (9). All these statistics were computed using the *"sjstats"* v 0.14 R package (10) and the code provided in (8).

Assessing the variance explained by a single SNP

In order to calculate the variance explained by the SNPs highlighted by our GWAS, we used the "PVE" method of Shim et al. (11), based on the effect sizes computed in the full regression model (which include the target SNP, genotype principal components, and other fixed-effect covariates). While this method does not explicitly account for the particularities of mixed model regression, as the procedure described above, it avoids the known problems of partitioning R^2 metrics between individual predictors (12). It also has the desirable property of accounting for the minor allele frequency of the SNP, thus avoiding overestimation of uncommon variants with larger effects.

Prioritising SNPs to be incorporated into the full regression model

The GWAS of norclozapine plasma concentrations and the clozapine/norclozapine metabolic ratio implicated several loci with complex association signals (**Results**). From the lists of credible SNPs generated by FINEMAP, we extracted the most credible missense SNP of each locus to incorporate into our full regression models (SNP + PCs + covariates). While in one of the cases this missense SNP was also the most credible SNP of the locus (norclozapine: rs2011425, **Supplementary Table S3**), SNPs with higher posterior probability existed in the other loci, though all were in high LD ($r^2 > 0.9$). We justify the selection of missense SNPs in these situations on the basis of their higher prior probability of driving GWAS associations, which has been well-stablished by other studies (13, 14). The parsimony of this procedure is also demonstrated by the fact that it results in the same UGT2B10 missense SNP being used in the models of both norclozapine and the metabolic ratio, where it shows the expected effect in opposite directions (**Table 2**). Finally, every missense SNPs incorporated in our regression models has been shown to impact protein functionality in direct experimental *in vitro* assays (15-17). Despite the convergence of these lines

of evidence, we caution that the confirmation of these SNPs as truly causal requires experimental validation and functional follow-up (18).

Polygenic score-based analysis of environmental predictors

Both smoking and weight have been shown to influence clozapine plasma concentrations (19). However, we cannot directly measure their relevance in our sample, as we lack the required clinical or self-report data. In order to assess the possibility of confounding effects from these exposures on our genetic associations, we carried out secondary GWAS controlling for polygenic risk scores for these variables. In order to generate the scores, we used the latest publicly available GWAS results for body mass index (BMI) (20) and cigarettes-per-day (21), as genetic markers associated to these traits have been shown to be predictive of related conditions, such as obesity (22) or nicotine dependence (23). The GWAS results were restricted to nominally significant LD-independent SNPs ($P < 0.05$; $r^2 < 0.1$), and the scores were calculated using PLINK v.1.9.

All the secondary GWAS carried out controlling for the smoking and BMI polygenic scores gave very similar results to those reported in the main text, with no gain or loss of genome-wide significant signals. Also, adding these polygenic scores to each of the models used to calculate genetic effects did not result in statistically different model fits, based on likelihood-ratio tests. Finally, presence of the scores in these models did not significantly change the effect size estimates of any of the GWAS-identified SNPs, which remained within one standard deviation of the original model. However, we found nominally significant associations of the scores with some of our outcomes, and thus we report their detailed statistics (**Supplementary Table S5**).

Given that our results showed similarities between clozapine and caffeine metabolism, we generated a polygenic score for the most likely environmental source of this xenobiotic, daily cups of coffee, based on recent GWAS results (24). For this we followed the approach of (25), which used only genome-wide significant SNPs, as such a score has been found to be correlated with coffee and tea consumption in UK Biobank data. However, since one of the SNPs used (rs2470893) is in strong LD (r^2=0.728) with our main GWAS signal for clozapine plasma concentration (rs2472297), we removed it from the computation to avoid collinearity. The resulting coffee polygenic score was found to be significantly associated with all of the plasma concentrations we assessed, even after including all other possible covariates (**Supplementary Table S5**). A regression model including all three polygenic scores described (BMI, tobacco and coffee) was shown, via likelihood-ratio tests, to fit the clozapine plasma levels data better than the model without

them (P = 0.001), though the effect size estimates of the rs2472297 SNP in the two models were within one standard deviation.

Analysis of gene expression data

In order to pinpoint putatively causal genes in the multigenic loci identified in the GWAS, we first assessed the gene expression pattern of the genes overlapping genome-wide-significant loci using the "GENE2FUNC" enrichment test implemented in FUMA v1.33 (26), with reference data from 53 tissue types from the GTEx project v7 (27). Within the clozapine levels analysis, the only significant result after multiple-testing correction was an under-expression in musculo-skeletal tissue (p_{CORR}=0.047). For norclozapine, we found significant over-expression of its associated genes in liver (p_{CORR}=1.35x10^{-6}), followed by salivary gland (p_{CORR}=9.75x10^{-6}) and esophagus mucosa (p_{CORR}=0.014). For the metabolic ratio, again we found significant over-expression in liver (p_{CORR}=8.41x10^{-9}) followed by terminal ileum (p_{CORR}=0.003) and stomach (p_{CORR}=0.007). Given the convergent evidence towards liver, and the fact that clozapine is known to be metabolised by liver enzymes, we focused the rest of our analysis on this tissue (28, 29). Using the TWAS approach implemented in FUSION (30) we queried the GTEx v7 liver data for association between our plasma concentration data and liver-specific gene expression. After Bonferroni correction for the number of hepatic cis-heritable genes detected (730), none of these analyses identified any significant genes in which the GWAS signal for any phenotype could be traced back to an alteration of gene expression.

FIGURE S1. Number of plasma concentration assays (time points) per individual in CLOZUK2

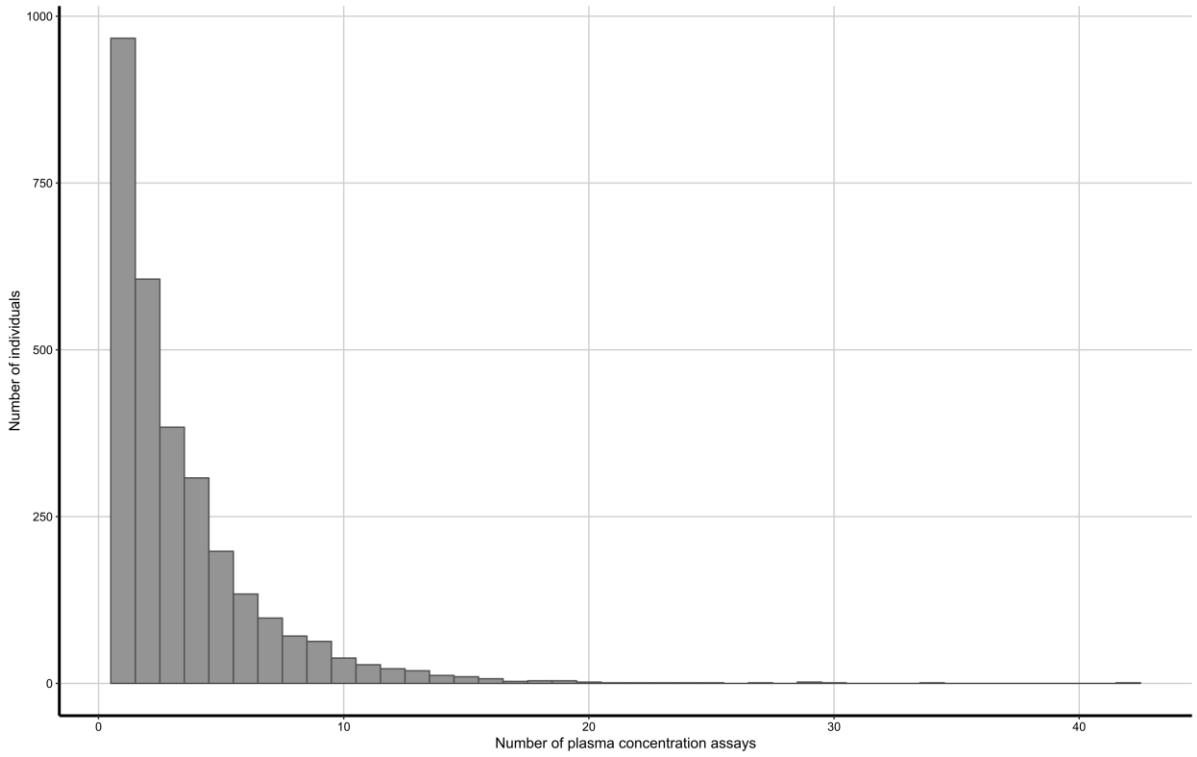


FIGURE S2. Plasma concentration distribution of clozapine (A) and norclozapine (B) in the CLOZUK2 sample

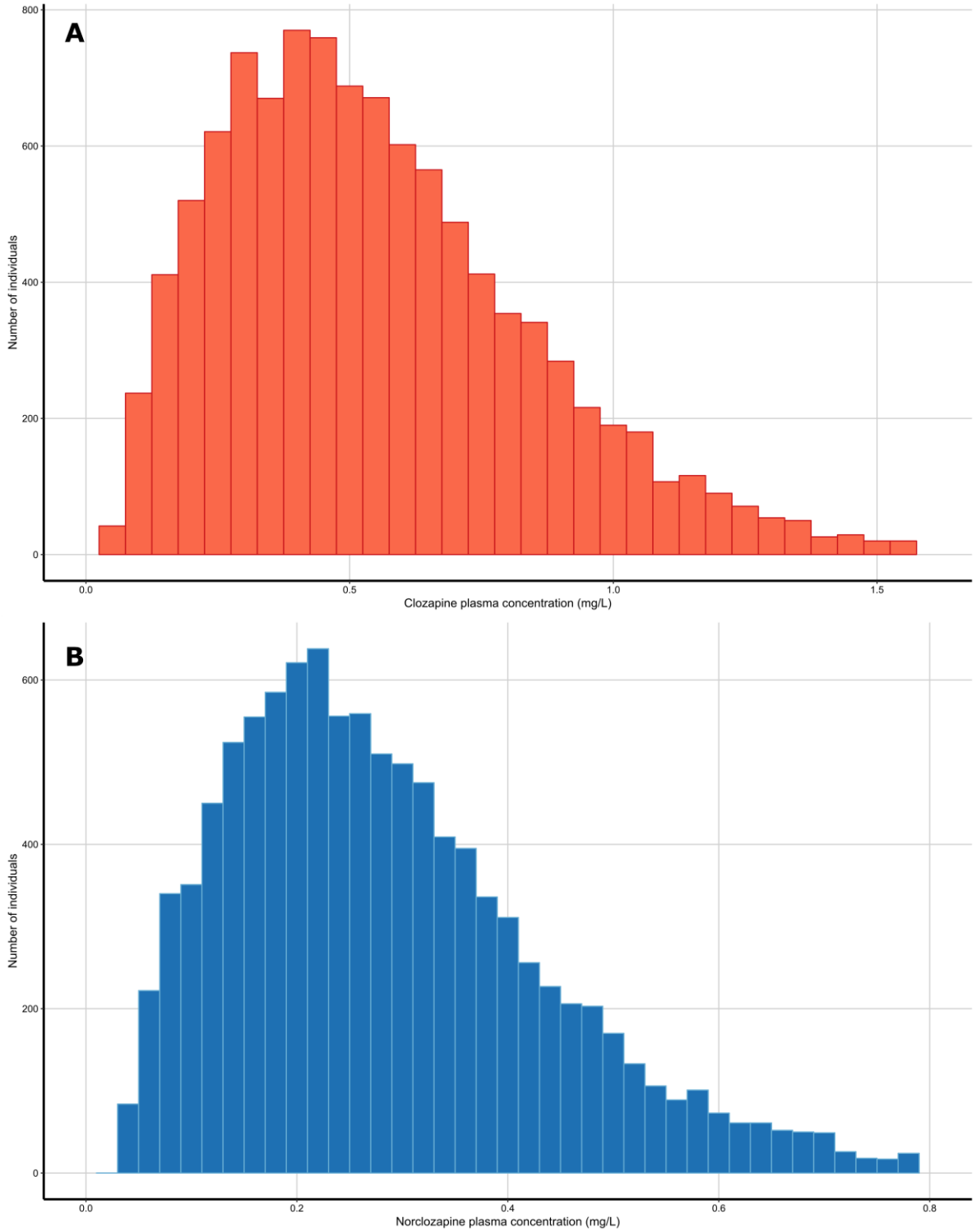


FIGURE S3. QQ plot of the clozapine plasma concentration GWAS

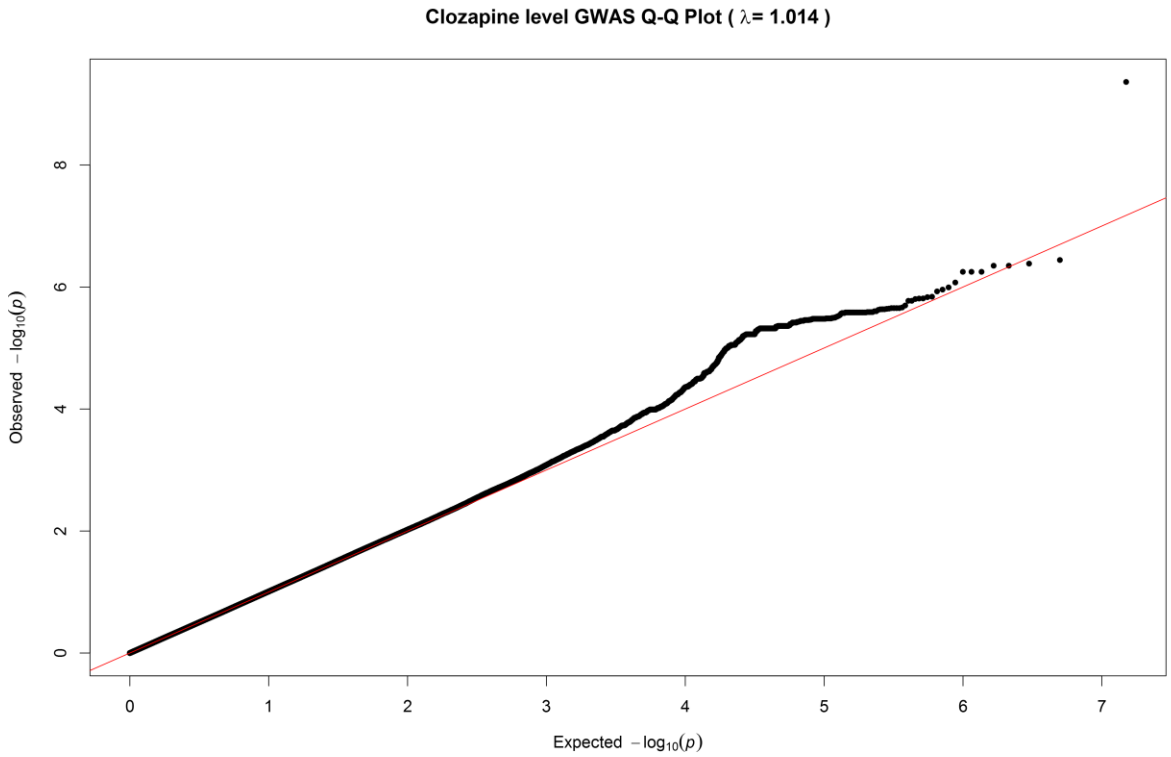


FIGURE S4. QQ plot of the norclozapine plasma concentration GWAS

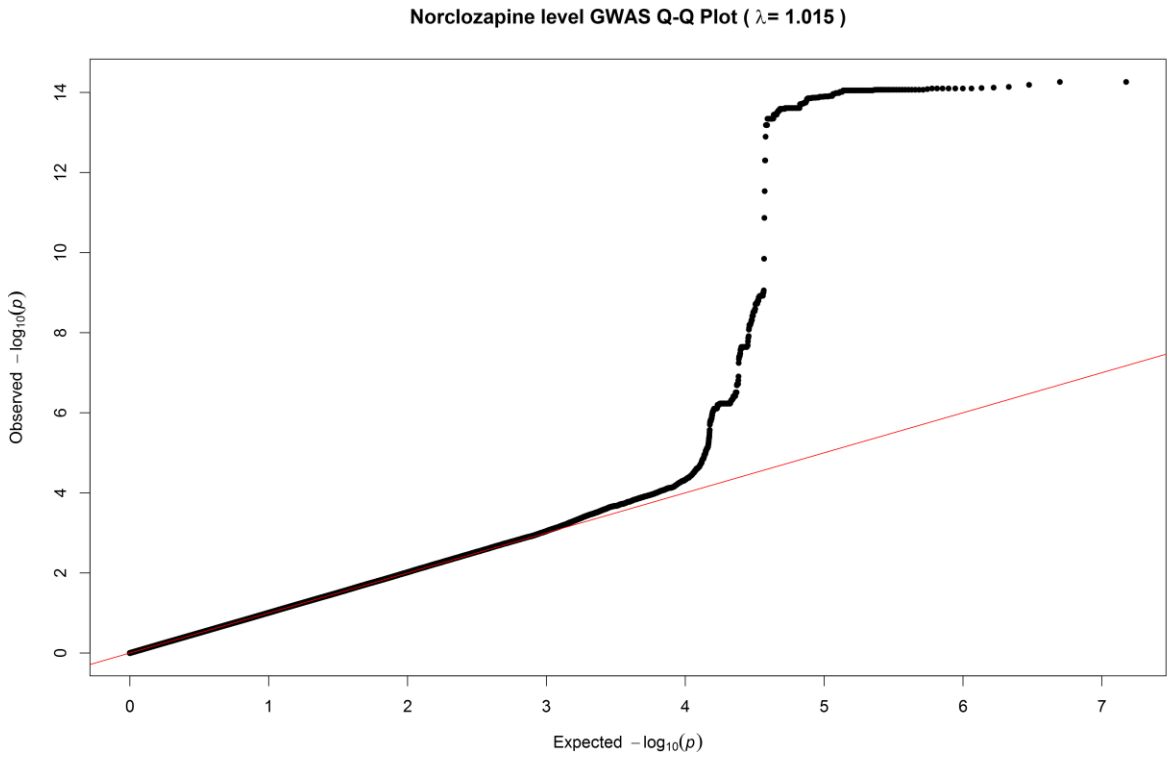


FIGURE S5. QQ plot of the clozapine/norclozapine metabolic ratio GWAS

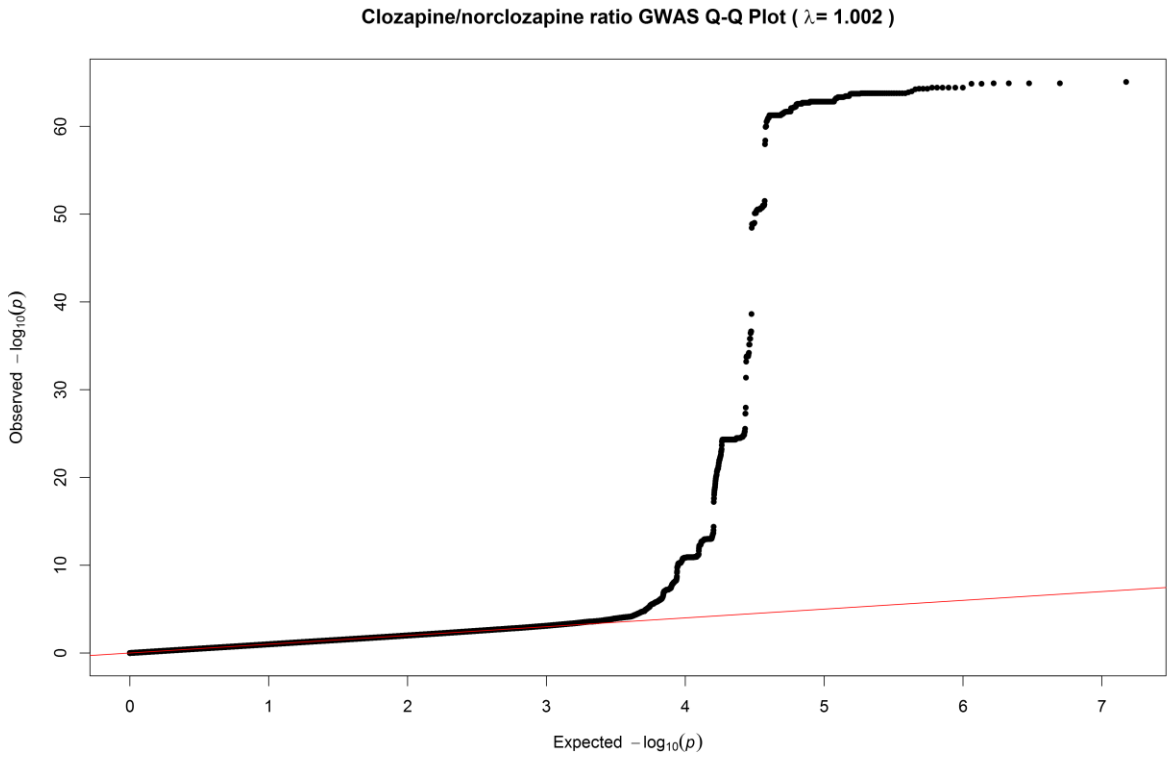


FIGURE S6. Manhattan plots from the analyses of mean clozapine (A), norclozapine (B) and metabolic ratio (C); and maximum clozapine (D), norclozapine (E) and metabolic ratio (F). Note the similarity with the results shown in Figure 1, albeit not all of the genome-wide significant loci can be observed as such.

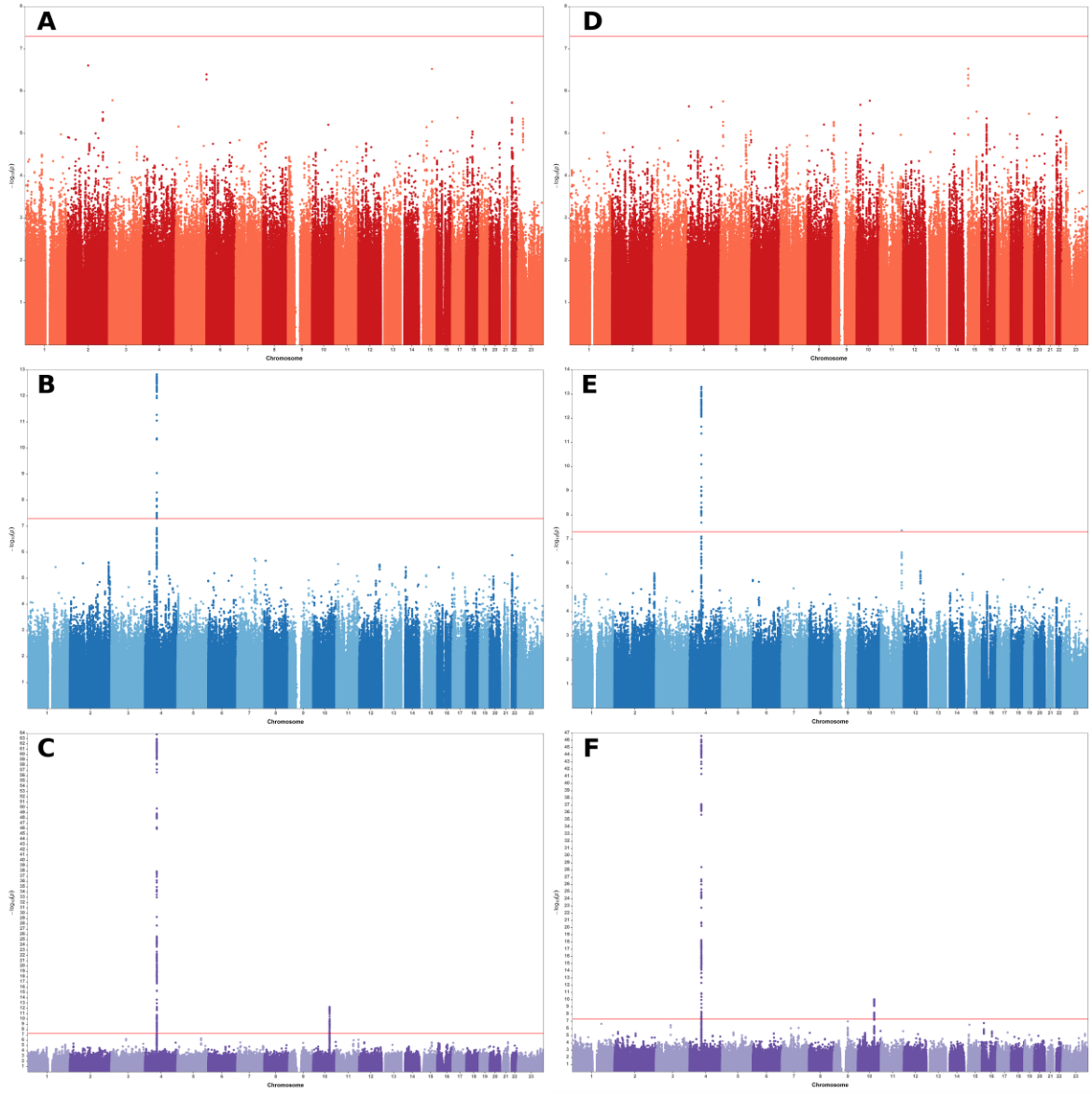


FIGURE S7. Distribution of effect sizes of independent SNPs tentatively associated ($p < 10^{-5}$) to clozapine metabolite plasma concentrations, across the minor allele frequency range. Bars are proportional to the standard error of the effect size estimate

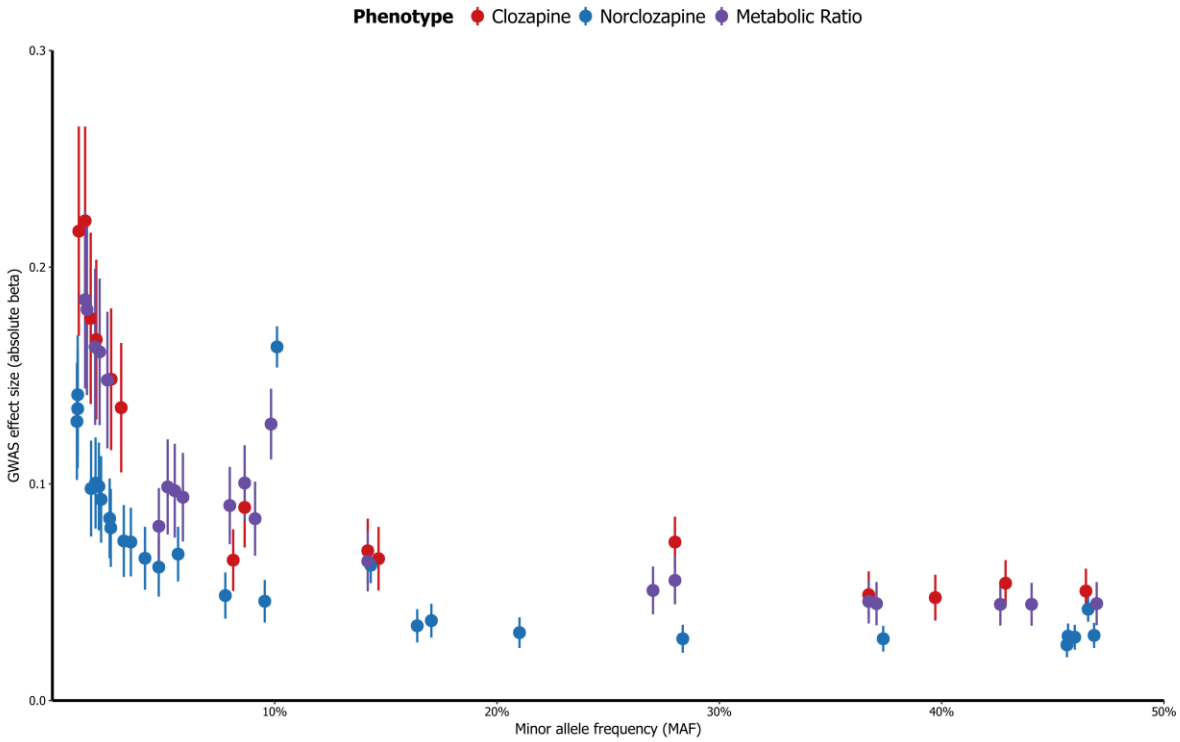
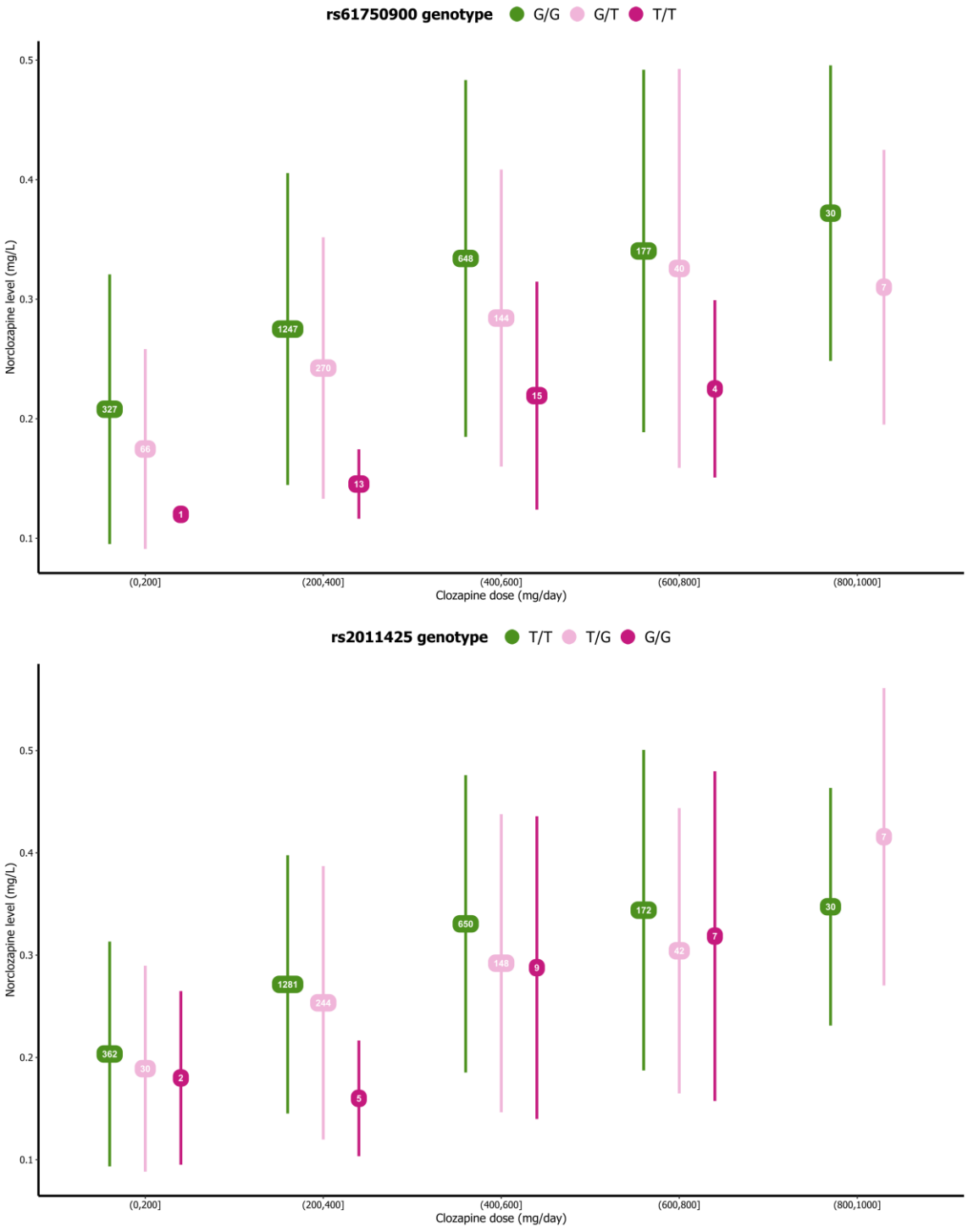


FIGURE S8. Effect of the norclozapine-associated genotypes on norclozapine plasma levels, at different daily clozapine doses. For this analysis, only the last time point of each CLOZUK2 individual was used. For each interval of daily clozapine dose, average norclozapine plasma concentrations and their standard deviations are shown. Values inside the central point represent the number of individuals within each genotype/interval category.



REFERENCES

1. Delignette-Muller ML, Dutang C: fitdistrplus: An R package for fitting distributions. J Stat Softw 2015;64:1-34.
2. Lindsey J, Jones B, Jarvis P: Some statistical issues in modelling pharmacokinetic data. Stat Med 2001;20:2775-2783.
3. 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.
4. Stephens MA: EDF statistics for goodness of fit and some comparisons. J Am Stat Assoc 1974;69:730-737.
5. Stasinopoulos DM, Rigby RA: Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. J Stat Softw 2007;23:46.
6. LaHuis DM, Hartman MJ, Hakoyama S, et al.: Explained variance measures for multilevel models. Organizational Research Methods 2014;17:433-451.
7. Nakagawa S, Schielzeth H: A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol 2013;4:133-142.
8. Nakagawa S, Johnson PC, Schielzeth H: The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. J R Soc Interface 2017;14:20170213.
9. Nakagawa S, Schielzeth H: Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biol Rev Camb Philos Soc 2010;85:935-956.
10. Lüdecke D: (2017). sjstats: Statistical Functions for Regression Models. Retrieved from https://CRAN.R-project.org/package=sjstats
11. Shim H, Chasman DI, Smith JD, et al.: A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. PLoS One 2015;10:e0120758.
12. Grömping U: Estimators of relative importance in linear regression based on variance decomposition. Am Stat 2007;61:139-147.
13. Pal LR, Moult J: Genetic Basis of Common Human Disease: Insight into the Role of Missense SNPs from Genome-Wide Association Studies. J Mol Biol 2015;427:2271-2289.
14. Sveinbjornsson G, Albrechtsen A, Zink F, et al.: Weighting sequence variants based on their annotation increases power of whole-genome association studies. Nat Genet 2016;48:314.
15. Kaminsky LS, de Morais SM, Faletto MB, et al.: Correlation of human cytochrome P4502C substrate specificities with primary structure: warfarin as a probe. Mol Pharmacol 1993;43:234.

16. 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.
17. Zhou J, Argikar UA, Remmel RP: Functional analysis of UGT1A4-P24T and UGT1A4-L48V variant enzymes. Pharmacogenomics 2011;12:1671-1679.
18. Visscher PM, Wray NR, Zhang Q, et al.: 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 2017;101:5-22.
19. 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.
20. Locke AE, Kahali B, Berndt SI, et al.: Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518:197-206.
21. Tobacco and Genetics Consortium: Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet 2010;42:441-447.
22. Belsky DW, Moffitt TE, Sugden K, et al.: Development and Evaluation of a Genetic Risk Score for Obesity. Biodemography Soc Biol 2013;59:10.1080/19485565.19482013.19774628.
23. Belsky DW, Moffitt TE, Baker TB, et al.: Polygenic risk accelerates the developmental progression to heavy, persistent smoking and nicotine dependence: Evidence from a 4-Decade Longitudinal Study. JAMA Psychiatry 2013;70:534-542.
24. Cornelis MC, Byrne EM, Esko T, et al.: Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption. Mol Psychiatry 2015;20:647-656.
25. Taylor AE, Smith GD, Munafo MR: Associations of coffee genetic risk scores with consumption of coffee, tea and other beverages in the UK Biobank. Addiction 2017;[in press].
26. Watanabe K, Taskesen E, van Bochoven A, et al.: Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8:1826.
27. 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.
28. Jann MW, Grimsley SR, Gray EC, et al.: Pharmacokinetics and Pharmacodynamics of Clozapine. Clin Pharmacokinet 1993;24:161-176.
29. Pirmohamed M, Williams D, Madden S, et al.: Metabolism and bioactivation of clozapine by human liver in vitro. J Pharmacol Exp Ther 1995;272:984-990.
30. Gusev A, Ko A, Shi H, et al.: Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 2016;48:245-252.