Published in final edited form as:

Biol Psychiatry. 2010 January 15; 67(2): 133–138. doi:10.1016/j.biopsych.2009.08.029.

A Genome-Wide Association Study of Citalopram Response in Major Depressive Disorder

Holly A. Garriock^{1,#}, Jeffrey B. Kraft^{1,#}, Stanley I. Shyn¹, Eric J. Peters¹, Jennifer S. Yokoyama¹, Gregory D. Jenkins², Megan S. Reinalda², Susan L. Slager², Patrick J. McGrath³, and Steven P. Hamilton^{1,*}

¹ Department of Psychiatry and Institute for Human Genetics, University of California, San Francisco, CA

² Department of Health Sciences Research, Mayo Clinic College of Medicine

³ Department of Psychiatry, Columbia University College of Physicians and Surgeons and New York State Psychiatric Institute

Abstract

Background—Antidepressant response is likely influenced by genetic constitution, but the actual genes involved have yet to be determined. We have carried out a genome-wide association study to determine if common DNA variation influences antidepressant response.

Methods—Our sample is derived from Level 1 participants in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, all treated with citalopram. Association for the response phenotype included 883 responders and 608 non-responders. For the remission phenotype, 743 subjects that achieved remission were compared to 608 non-responders. We used a subset of SNPs ($n = 430,198$) from the Affymetrix 500K and 5.0 Human SNP Arrays, and association analysis was carried out after correcting for population stratification.

Results—We identified three SNPs associated with response with p -values less than 1×10^{-5} near the *UBE3C* gene (rs6966038, $p = 4.65 \times 10^{-7}$), another 100kb away from *BMP7* (rs6127921, $p = 3.45 \times 10^{-6}$), and a third that is intronic in the *RORA* gene (rs809736, $p = 8.19 \times 10^{-6}$). These same SNPs were also associated with remission. Thirty-nine additional SNPs are of interest with p -values ≤ 0.0001 for the response and remission phenotypes.

Conclusions—Although the findings reported here do not meet a genome-wide threshold for significance, the regions identified from this study provide targets for independent replication and novel pathways to investigate mechanisms of antidepressant response. This study was not placebo controlled, making it possible that we are also observing associations to non-specific aspects of drug treatment of depression.

Keywords

genetics; citalopram; STAR*D; ARNTL; ACCN1; UBE3C; antidepressant

*corresponding author Steven P. Hamilton, MD, PhD, Carol Cochran Schaffner Endowed Chair in Mental Health, University of California, San Francisco, Department of Psychiatry, 401 Parnassus Ave, Box NGL-0984, San Francisco, CA, 94143-0984, steveh@lppi.ucsf.edu.

#these authors contributed equally to this work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

INTRODUCTION

Depression is a serious health concern with a lifetime prevalence of ~16% in the United States (1). Symptoms of Major Depressive Disorder (MDD) are often alleviated with use of antidepressant medication, but about 65% of those afflicted with MDD will not achieve remission after their first medication trial (2). Identification of patient factors that increase the likelihood of successful treatment would benefit patients, clinicians, and society. Such factors may include inter-individual genetic differences that contribute to variability in antidepressant response. It has been observed that antidepressant response clusters in families, supporting a role for genetic variation in antidepressant response, though the degree of heritability is still unknown (3,4). Previously, we reported candidate gene association studies for pharmacokinetic genes (5) and serotonin-pathway genes (6,7) in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) sample, which is comprised of 1,953 subjects treated with citalopram. Although we failed to find statistically significant associations between antidepressant response and these candidate genes, other interesting candidate genes have been reported for response (8,9) and remission (10,11) in the same sample. Given the lack of placebo in the STAR*D trial, it is possible that previous associations may be related to non-specific aspects of the treatment protocol.

An alternative to candidate gene studies are genome-wide association studies (GWAS), which have successfully identified genetic variants for numerous diseases, including type 2 diabetes (12) and Crohn's Disease (13). Several pharmacogenetic GWASs have also identified novel candidate regions for phenotypes of interest including hydrochlorothiazide response for treatment of hypertension (14), anti-tumor necrosis factor response for rheumatoid arthritis (15), interferon beta therapy for multiple sclerosis (16) and iloperidone response for schizophrenia (17). Using the STAR*D sample, we carried out a GWAS to determine if common DNA variation influences citalopram response and remission.

METHODS AND MATERIALS

Study Information

Subjects—The design of the Sequenced Treatments Alternatives to Relieve Depression (STAR*D) study has been previously described (2,18,19). To maximize power to detect associations, we focused on Level 1 data collected in the STAR*D trial, where all participants were initially treated with citalopram. Our DNA samples and phenotypes have been described in previous reports (5,20). For this report, our outcome phenotypes were antidepressant response and remission, the definitions of which have been previously described (5). Briefly, the first comparison is between responders and non-responders. Both groups required a 42-day enrollment in the study with responders demonstrating a $\geq 50\%$ reduction in the 16-item Quick Inventory of Depressive Symptomatology, self-report version (QIDS-SR) (21) from baseline to final visit. Non-responders did not meet this criterion. We use “response” to denote drug efficacy. For remission, we compared non-responders as described above to individuals scoring ≤ 5 on the QIDS-SR at follow-up, which closely corresponds with the conventional definition of remission by the Hamilton Rating Scale for Depression (HAM-D) ≤ 7 . Figure S1 details the inclusion and exclusion of subjects for the association study. Tolerance was based on exit data for the 1st level of the STAR*D study. If a subject continued on citalopram or entered follow-up at the end of level 1, they were considered tolerant to citalopram. If the level-exit or study-exit forms indicated side-effects, they were considered intolerant. Figure S1 details the inclusion and exclusion of subjects for the association study. For secondary analysis, response was considered as a quantitative variable and we measured the percent change in QIDS-SR from baseline.

Genotyping and Quality Control Procedures

Genotyping of the 1,948 STAR*D samples was conducted at two locations and on two different platforms. A total of 969 subjects were genotyped at Affymetrix, Inc. (South San Francisco) on the Human Mapping 500K Array Set. We genotyped the remaining 979 samples using the Affymetrix Genome-Wide Human SNP Array 5.0. The two groups were balanced by ethnic grouping, gender and proportions of responders and non-responders. Samples run on Affymetrix 500K Array were called using the BRLMM algorithm, and samples analyzed on Affymetrix Array 5.0 were called using the BRLMM-P algorithm. Twelve samples were genotyped on both the 500K and 5.0 Arrays, and we found > 99% concordance across these platforms.

Hardy-Weinberg Equilibrium was evaluated using either Fisher's exact test or chi-square test within each ethnic group, as appropriate. SNPs with $P < 1.0 \times 10^{-5}$ in all three ethnic categories (African-American, Non-Hispanic Caucasian, and Hispanic Caucasian) were excluded ($n=1,233$ SNPs). This was done to remove the poorly performing SNPs that could be a result of genotyping error. The requirement for the SNP to violate HWE in all three ethnic groups derived from the assumption that genotyping error would not be specific to one ethnic group. SNPs with minor allele frequencies $< 1.0\%$ ($n=10,792$) or call rate $< 95\%$ ($n=42,908$) were excluded. The final number of SNPs for autosomal analysis was 430,198. Differential missingness between responders and non-responders was assessed, and considered to not influence the association statistics. Table S13 lists the 63 SNPs with differential missingness p -values $< 1.0 \times 10^{-3}$. Only one SNP (rs2443527, rank 8,431) fell within the top 10,000 association hits for remission, therefore, no SNPs were dropped due to differential missingness. SNPs on the X-chromosome ($N=9,697$) that passed the same quality control measures as the autosomal SNPs were also tested for association with response and remission. The estimated average rate of coverage of common variants across the genome is 46%. This is based on an $r^2 \geq 0.8$ between tagging SNPs and non-panel HapMap SNPs (release 21) in the 10 ENCODE regions (22). We estimated coverage using of these regions because they are neutral with respect to choice of location and density of the SNPs on the Affymetrix array we used for our association study.

Individuals were excluded if they were of self-reported Native American ancestry ($n=16$), Asian ancestry ($n=21$), Pacific Island ancestry ($n=8$), mixed ancestry ($n=68$), or unknown ($n=1$) due to small numbers in these groups. Consequently, for the response phenotype, we analyzed 1,067 Non-Hispanic Caucasians (71.5%), 241 African-Americans (16.2%) and 183 Hispanic Caucasians (12.3%). Further, we excluded individuals if they had more than 5% of their genotypes missing ($n=71$), did not have follow-up phenotypic data ($n=16$), or were cryptically related to others in the sample ($n=32$), which was determined using the 1-IBS metric tabulated in PLINK v1.04 (22). The final number of individuals in the analysis was 1,491 for response vs. non-responders, and 1,351 for remitters vs. non-responders.

Assessment of Population Heterogeneity

Ancestry was controlled for using multidimensional scaling (MDS), which is a multivariate method that summarizes all SNPs into linear combinations of the count of rare alleles of SNPs. These MDS dimensions are linearly independent of each other, and are given consecutive numbering (i.e. MDS #1, MDS #2, etc.) based on how much variation in SNP rare allele count they explain. We used 140,701 independent SNPs in the MDS analysis. Independence was determined using the `-indep` function in PLINK for the following conditions: 50 SNPs in the window, five SNP window shift, and a variance inflation factor of two. The first ten MDS dimensions were then used as covariates in association analyses to adjust for population stratification, as previously established (23). Q-Q plots of expected and observed p -values are available for response and remission in Figures S2 and S3, and were created using publicly

available scripts (24,25). The inflation factor λ was computed based on the Cochran-Armitage trend tests.

Analysis

All SNPs were tested for association with response and remission using logistic regression. The minor allele homozygous genotype was used as the reference group, and each SNP was modeled in individuals as having a log-additive effect, after adjusting for the ten MDS dimensions. Odds ratios (ORs) and 95% confidence intervals (CIs) are reported. Statistical analysis was done using PLINK v1.04 (26). Cluster plots for all SNPs contributing to results reported in Tables 1 and 2 were examined for artifacts (see Figure S5).

We estimated statistical power using Quanto (27). We used the following as parameters: $\alpha = 1.16 \times 10^{-7}$ (Bonferroni correction for 430,198 markers), risk of response (0.59) and remission (0.55), $1-\beta = 0.8$, log additive model, and minor allele frequency of 0.25. For the response phenotype (883 responders and 608 non-responders), the genotypic relative risk that could be detected at the set power level was 1.74. For the remission phenotype (743 remitters and 608 non-responders), the equivalent detectable genotypic relative risk was 1.77. A range of power estimations for both phenotypes is graphically depicted in Figure S6.

Genotype and phenotype data are available from the NIMH Center for Collaborative Genetic Studies on Mental Disorders.

RESULTS

Demographics

We evaluated 430,198 SNPs for association with antidepressant response and remission. For response, 1,491 (608 non-responders; 883 responders) subjects were analyzed, and for remission, 1,351 (608 non-responders; 743 responders) subjects were analyzed. The gender distribution between the groups within each analysis was not significantly different (Table S1). There were significant differences by ethnicity between responders and non-responders ($p = 0.003$) and between remitters and non-responders ($p = 0.0003$). African Americans were over-represented among non-responders (19.7%), compared to responders (13.7%) and remitters (12.5%). Significant differences among responders and non-responders also existed for a number of demographic and clinical measures (e.g., insurance and employment status, clinical co-morbidities, etc), as shown in Table S1. Since these variables are correlated with the response phenotype, they are considered potential confounding variables. Therefore, we tested association of the top 100 SNPs associated with response for association with insurance status, employment status, years of completed schooling, history of attempted suicide, and co-morbid disorders (generalized anxiety, hypochondriasis, panic disorder, somatoform disorder, social phobia, and obsessive compulsive disorder). None of these variables are associated with the top reported SNPs, and therefore are not considered confounding variables for the response and remission association results, as confounding variables must be associated with outcome (e.g. response) and the genetic variant (e.g. our list of top 100 SNPs).

We used the same approach for assessment of tolerance as a confounding variable. The top 100 SNPs from the response and remission association tests were tested for association with tolerance. This resulted in 145 SNPs be tested for association with tolerance, as there is overlap in the association results between response and remission. Seventeen SNPs were found to have association p-values < 0.05 with tolerance, with the smallest p-value = 0.0003, and are presented in Table S2. Only one SNP (rs11128623) met a Bonferroni corrected p-value ($p = 0.00034$) for significance given that 145 SNPs were evaluated, and thus could be considered as confounded by tolerance.

Association Results

Q-Q plots and full genome plots for the association tests of both the response and remission phenotypes are found in supplemental data (Figures S2, S3 and S7). The genomic inflation factors for both phenotype association tests are 1.01, suggesting that our association results are unlikely to be confounded by population stratification. Figure S7 shows regional plots for the two strongest association results for response and remission. We found 39 SNPs and 41 SNPs with p -values $< 1.0 \times 10^{-4}$ for the response and remission phenotypes, respectively. Table 1 lists the top 25 findings for the response phenotype and Table 2 lists the top findings for the remission phenotype. A more comprehensive list of the top 100 results for both phenotypes is presented in Tables S3 and S4. Thirteen of the top 25 SNPs from results for the response phenotype analysis are also found in the top 25 findings for the remission phenotype. This is not surprising given the correlation between the two phenotypes. The top two results for response and remission are the same SNPs: rs6966038 and rs6127921, with p -values for response of 4.65×10^{-7} and 3.45×10^{-6} , and for remission, 3.63×10^{-7} and 1.07×10^{-6} . We sought to determine if additional SNPs in these regions of interest also showed evidence for association (Tables S5 and S6). Our top-ranked finding for both phenotypes does not have any additional SNPs with a p -value < 0.05 and an $r^2 \geq 0.50$ within 250kb. However, rs6127921 (rank 2) has a nearby SNP with a p -value < 0.001 . We analyzed the surrounding regions for the most associated SNPs for proxies in the region, using the --proxy function in PLINK. For the top SNP finding (for both response and remission), we found no proxies for this. However, the omnibus haplotype test supported association in the region ($p = 1.19 \times 10^{-5}$ for response, and 3.92×10^{-7} for remission). For the second ranked SNP associated with response, we found an omnibus haplotype test $p = 0.00065$, with 2 of 6 subhaplotypes meeting proxy criteria with p -values 0.00036 and 0.00046. These analyses suggest although no single SNPs show similar associations to the most associated SNPs in these regions, we found haplotypes that do show association results of a similar magnitude and that are correlated with the original associated SNPs. The top 10 association results for response and remission on the X chromosome and pseudo-autosomal regions are in Table S7. No statistically significant association was detected with these phenotypes on the sex chromosomes.

The ancestry groups were also analyzed separately for the response phenotype. The top 25 hits for each ancestry group (Non-Hispanic Caucasians, Hispanic Caucasians, and Non-Hispanic African Americans) are presented in Table S8. There is no correlation of the SNP rankings between the ancestry groups, most likely due to low power to detect risk variants in the two smaller sub-samples, or differential risk loci related to ethnicity. In an additional exploratory analysis where response was considered as a quantitative trait (percent change in QIDS-SR from baseline), the lowest p -value was 3.20×10^{-6} for rs6508329. The top 100 association p -values for this analysis are presented in Table S9.

We found high correlation between SNP ranking in association results when analyzed with 1 multi-dimensional scaling (MDS) covariate vs. 10 MDS covariates ($r^2 = 0.974$). Table S10 demonstrates that whether we correct for 1 or 10 covariates, the results are largely the same. Further, as an alternative to SNP ranking we also evaluated change in odds ratios (ORs) and standard deviations and found little change across these comparisons (e.g., mean difference in ORs = 0.00013), suggesting they are comparable metrics.

Analysis of Prior Candidate Genes

Many candidate gene studies of antidepressant response have preceded this current GWAS, some with positive and many with null associations. We assessed our data for how these past candidates performed using WGAviewer (28). A list of the investigated genes, and their corresponding ranks and p -values from the GWAS data are listed for the response and remission phenotypes (Table S11). Additionally, the gene coverage [coverage rate = (tagged SNPs in

HapMap + genotyped SNPs in GWAS set)/total common SNPs in HapMap; an algorithm created in WGA viewer (28,29)], based on our SNP set, and numbers of SNPs with p-value < 0.05 are listed in Table S12. The serotonin receptor 2A (*HTR2A*) has several nominally associated SNPs for response and remission. This supports previous findings of *HTR2A* being involved in antidepressant response using the same STAR*D sample used here (7,9). The D2 dopamine receptor (*DRD2*) shows some evidence of association for one SNP (rs2234689, p-value = 0.0001 for remission). Notably, no support for the serotonin transporter (*SLC6A4*) was detected in remission or response analyses. This finding also supports our previous report of no association in this gene in the same sample (6).

DISCUSSION

Our findings represent a GWAS conducted to determine if genetic variation influences population variation in antidepressant response and remission in Major Depressive Disorder in the STAR*D sample. Our single locus findings for response and remission include two SNPs located between two sets of genes. The top finding (rs6966038; response $p = 1.65 \times 10^{-7}$) is 51kb from the ubiquitin protein ligase E3C gene (*UBE3C*) and 77kb from the motor neuron and pancreas homeobox 1 (*MNX1* or *HLXB9*) gene. The UBE3C protein modifies proteins to signal them for degradation (30) and the MNX1 protein is a DNA binding protein that controls gene expression at the caudal end of the developing notochord of an embryo (31). The second finding (rs6127921; response $p = 3.45 \times 10^{-6}$) is closest to the bone morphogenic protein 7 (*BMP7*) gene, but at least 106kb away. Many of the SNPs that appear in the top tier of results in the response phenotype also appear in the remission phenotype. This is likely due to the overlap in samples and phenotypic definition. In our case, all remitters are also responders, however, not all responders achieve remission, and thus remitters are a subgroup of the responders. Furthermore, both phenotypes were compared to a common non-responder group.

When we examined the local regions around our highest ranked findings, we found four gene regions supporting association to response and remission: amiloride-sensitive cation channel 1, neuronal (*ACCNI*), PR domain-containing protein 2 (*PRDM2*), aryl hydrocarbon receptor translocator-like (*ARNTL*) and nucleolar protein 4 (*NOLA*) genes. More detailed information regarding these genes is available in the Supplemental Materials.

Although these data potentially implicate novel pathways for investigation of mechanisms of antidepressant response, we do not believe that these data can reliably be incorporated into genetic tests for prediction of antidepressant response. Rigorous replication in a larger, independent sample demonstrating significant signals of similar effect size and direction would be required before a solid correlation between these variants and antidepressant response could be surmised. The effect sizes for our highest ranked SNPs are quite low and, though interesting, do not justify a clinical application at this early stage of investigation. While this study represents the largest pharmacogenetic study of antidepressant response, our sample was only large enough to detect genotypic relative risks in the range of 1.75 or greater, and we observed several findings with odds ratios in that range. Given the limited power in our sample, meta-analysis of all available samples with similar pharmacogenetic phenotypes will be required before any definitive conclusions can be generated. Similarly, while our data could be construed as not supporting the role of common genetic variation in citalopram response, this conclusion may be premature from a single study. It is possible that larger samples may detect true risk alleles of small effect, but still of scientific interest. It is also possible that higher density genotyping of common alleles not covered by our genotyping platform would uncover missed associations.

Imputation of these data was not performed due to the admixed and multi-ethnic nature of our sample. The choice of a reference panel on which to base the imputation is not straightforward

and complicates an imputation procedure. Imputation would be useful to show regional support for a SNP that exceeds typical genome-wide statistical significance values; however we are not reporting that any of our SNPs surpass this threshold. Likewise, this report does not include copy number variation (CNV) analyses due to the technical challenges of calling CNVs in Affymetrix data when using two separate genotyping platforms and genotyped at two locations. Furthermore, the power to detect an association of CNVs with antidepressant response would be marginal given their rarity.

The limitations to our study design have been previously reviewed (5,6,56). Briefly, they include: phenotypic definitions, STAR*D study design for placebo and adherence, lack of generalizability to the general affected population and to other antidepressant medications, and population stratification. Each is discussed in turn.

Our phenotypes are antidepressant response and remission, and require a six week exposure to citalopram to be considered for analysis. This should theoretically increase our chances of finding a true drug response if we are excluding subjects who responded to treatment before the six week timeline; though it is possible that we are missing non-responders that dropped out of the study early. In a post-hoc analysis, we removed the six week treatment criterion, and found that inclusion of additional individuals did not improve the strength of our findings (Table S14 and Figure S8). Another concern is that the STAR*D trial did not have a placebo arm in the clinical trial design, and did not measure serum levels or other levels of adherence to treatment. Under these conditions we cannot thoroughly assess true response to treatment versus placebo or be sure that all responders are responding due to the medication. This is a situation shared with all treatment studies in depression which attempt to discover predictors of response are faced with, namely there is still no agreed upon method to separate true drug response from placebo response. Thus, in the absence of placebo or comparator in Level 1, it is possible that we are not studying citalopram response, per se, but that our observations may actually relate to the speed of recovery or the placebo effect. This intriguing possibility provides grounds for further research. Likewise, and more relevant is that we cannot know for sure that non-responders were actually taking their medication, and might be lacking a response due to non-compliance. Furthermore, we lack information on concomitant medications that could interact with citalopram and alter its effects (positively or negatively) and lead to an artificial inflation of the phenotype (response or non-response). Although this is a real issue in this design, STAR*D response rates are similar to clinical trials in which adherence was measured, suggesting that the rates observed in the STAR*D sample may be fairly accurate.

Another limitation of the STAR*D study design for genome wide association studies is the initiation of DNA collection after the start of the study. As a result, not all individuals in the STAR*D study donated DNA, and there are some differences between individuals who donated DNA and those who did not, potentially reducing the generalizability of our results to the entire STAR*D sample. These differences are outlined in previous reports (6) and include variables such as education, marital status, age, ethnicity and severity of disease.

The STAR*D study design was intended to mimic naturalistic clinical practices as closely as possible and allowed dosage adjustments as necessary, presenting a possible confound for a GWAS on antidepressant response phenotypes. We found this not to be the case, since there was no significant difference between responders and non-responders for final citalopram dose. Our findings may be specific to citalopram treatment, and may not be the same for treatment with other selective serotonin reuptake inhibitors (SSRI's) or other classes of antidepressant medication. Notably, this is the largest collection of samples designed to study antidepressant response to date. Until another study of similar magnitude is conducted with a different SSRI or another class of treatment, we are unable to determine if the findings reported here are relevant for response to medications other than citalopram.

It is possible that for very few of the SNPs, the response or remission association results could be confounded by tolerance. However, only one SNP showed possible evidence of confounding, suggesting that confounding from tolerance is not influencing our results.

Population stratification is a concern in genetic association studies, regardless of phenotype. We controlled for population structure by using MDS vectors as covariates in our logistic regression. This approach has been shown to be effective in accounting for confounding due to population structure (23), and is additionally supported by our λ estimates of 1.01 (for both response and remission), which also strongly suggests limited confounding due to hidden population structure in our data.

In summary, we report results of the first genome wide association study on antidepressant response phenotypes in the STAR*D sample. These data will prove useful for meta-analysis of similar antidepressant GWAS datasets. Additionally, they provide a basis for several regions or specific genes to be pursued in fine-mapping and independent replication studies. Furthermore, these data invite investigation of the associated proteins in novel mechanisms of antidepressant response.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institute of Mental Health (NIMH) to S.P.H. (MH072802); by NIMH training funds to H.A.G. (F32 MH082562-01A1 & T32 MH19552-11), and to S.I.S. (R25 MH060482-08 & T32 MH19126-19); by NARSAD New Investigator Award A109584 (H.A.G.); by a Howard Hughes Medical Institute Pre-Doctoral Fellowship to E.J.P.; and by the State of New York for partial support to P.J.M. for this work. The authors appreciate the efforts of the STAR*D Investigator Team for acquiring, compiling, and sharing the STAR*D clinical dataset. STAR*D was funded by NIMH via contract (N01MH90003) to the University of Texas Southwestern Medical Center at Dallas (A. John Rush, P.I.). We would also like to thank Stephen Wisniewski, Ph.D., Director, STAR*D Data Coordinating Center, University of Pittsburgh for providing demographic and clinical response data.

FINANCIAL DISCLOSURES

Dr. McGrath reports receiving research support from the National Institute of Mental Health, the New York State Department of Mental Hygiene, and Roche Pharmaceuticals. The other authors reported no biomedical financial interests or potential conflicts of interest.

References

1. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003;289:3095–3105. [PubMed: 12813115]
2. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 2006;163:28–40. [PubMed: 16390886]
3. Franchini L, Serretti A, Gasperini M, Smeraldi E. Familial concordance of fluvoxamine response as a tool for differentiating mood disorder pedigrees. *J Psychiatr Res* 1998;32:255–259. [PubMed: 9789202]
4. O'Reilly RL, Bogue L, Singh SM. Pharmacogenetic response to antidepressants in a multicase family with affective disorder. *Biol Psychiatry* 1994;36:467–471. [PubMed: 7811843]
5. Peters EJ, Slager SL, Kraft JB, Jenkins GD, Reinalda MS, McGrath PJ, et al. Pharmacokinetic genes do not influence response or tolerance to citalopram in the STAR*D sample. *PLoS ONE* 2008;3:e1872. [PubMed: 18382661]

6. Kraft JB, Peters EJ, Slager SL, Jenkins GD, Reinalda MS, McGrath PJ, et al. Analysis of association between the serotonin transporter and antidepressant response in a large clinical sample. *Biol Psychiatry* 2007;61:734–742. [PubMed: 17123473]
7. Peters EJ, Slager SL, Jenkins GD, Reinalda MS, Garriock HA, Shyn SI, et al. Resequencing of serotonin-related genes and association of tagging SNPs to citalopram response. *Pharmacogenet Genomics* 2009;19:1–10. [PubMed: 19077664]
8. Paddock S, Laje G, Charney D, Rush AJ, Wilson AF, Sorant AJ, et al. Association of GRIK4 with outcome of antidepressant treatment in the STAR*D cohort. *Am J Psychiatry* 2007;164:1181–1188. [PubMed: 17671280]
9. McMahon FJ, Buervenich S, Charney D, Lipsky R, Rush AJ, Wilson AF, et al. Variation in the gene encoding the serotonin 2A receptor is associated with outcome of antidepressant treatment. *Am J Hum Genet* 2006;78:804–814. [PubMed: 16642436]
10. Lekman M, Laje G, Charney D, Rush AJ, Wilson AF, Sorant AJ, et al. The FKBP5-gene in depression and treatment response--an association study in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Cohort. *Biol Psychiatry* 2008;63:1103–1110. [PubMed: 18191112]
11. Perlis RH, Moorjani P, Fagerness J, Purcell S, Trivedi MH, Fava M, et al. Pharmacogenetic Analysis of Genes Implicated in Rodent Models of Antidepressant Response: Association of TREK1 and Treatment Resistance in the STAR(*)D Study. *Neuropsychopharmacology*. 2008
12. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008;40:638–645. [PubMed: 18372903]
13. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008;40:955–962. [PubMed: 18587394]
14. Turner ST, Bailey KR, Fridley BL, Chapman AB, Schwartz GL, Chai HS, et al. Genomic association analysis suggests chromosome 12 locus influencing antihypertensive response to thiazide diuretic. *Hypertension* 2008;52:359–365. [PubMed: 18591461]
15. Liu C, Batliwalla F, Li W, Lee A, Roubenoff R, Beckman E, et al. Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis. *Mol Med* 2008;14:575–581. [PubMed: 18615156]
16. Byun E, Caillier SJ, Montalban X, Villoslada P, Fernandez O, Brassat D, et al. Genome-wide pharmacogenomic analysis of the response to interferon beta therapy in multiple sclerosis. *Arch Neurol* 2008;65:337–344. [PubMed: 18195134]
17. Lavedan C, Licamele L, Volpi S, Hamilton J, Heaton C, Mack K, et al. Association of the NPAS3 gene and five other loci with response to the antipsychotic iloperidone identified in a whole genome association study. *Mol Psychiatry*. 2008
18. Fava M, Rush AJ, Trivedi MH, Nierenberg AA, Thase ME, Sackeim HA, et al. Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatr Clin North Am* 2003;26:457–494. x. [PubMed: 12778843]
19. Rush AJ, Trivedi M, Fava M. Depression, IV: STAR*D treatment trial for depression. *Am J Psychiatry* 2003;160:237. [PubMed: 12562566]
20. Kraft JB, Slager SL, McGrath PJ, Hamilton SP. Sequence analysis of the serotonin transporter and associations with antidepressant response. *Biol Psychiatry* 2005;58:374–381. [PubMed: 15993855]
21. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 2003;54:573–583. [PubMed: 12946886]
22. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007;447:799–816. [PubMed: 17571346]
23. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–909. [PubMed: 16862161]

24. Saxena R, Voight BF, Lyssenko V, Burt NP, de Bakker PI, Chen H, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;316:1331–1336. [PubMed: 17463246]
25. Broad Institute Plotting Genome-Wide Association Results
26. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575. [PubMed: 17701901]
27. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol* 2002;155:478–484. [PubMed: 11867360]
28. Ge D, Zhang K, Need AC, Martin O, Fellay J, Urban TJ, et al. WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res* 2008;18:640–643. [PubMed: 18256235]
29. Ge, D.; Goldstein, DB. WGAViewer. <http://people.genome.duke.edu/~dg48/WGAViewer/>
30. Urso ML, Chen YW, Scrimgeour AG, Lee PC, Lee KF, Clarkson PM. Alterations in mRNA expression and protein products following spinal cord injury in humans. *J Physiol* 2007;579:877–892. [PubMed: 17218363]
31. Ross AJ, Ruiz-Perez V, Wang Y, Hagan DM, Scherer S, Lynch SA, et al. A homeobox gene, HLXB9, is the major locus for dominantly inherited sacral agenesis. *Nat Genet* 1998;20:358–361. [PubMed: 9843207]
32. Waldmann R, Champigny G, Voilley N, Lauritzen I, Lazdunski M. The mammalian degenerin MDEG, an amiloride-sensitive cation channel activated by mutations causing neurodegeneration in *Caenorhabditis elegans*. *J Biol Chem* 1996;271:10433–10436. [PubMed: 8631835]
33. Price MP, Lewin GR, McIlwrath SL, Cheng C, Xie J, Heppenstall PA, et al. The mammalian sodium channel BNC1 is required for normal touch sensation. *Nature* 2000;407:1007–1011. [PubMed: 11069180]
34. Stone JL, Merriman B, Cantor RM, Geschwind DH, Nelson SF. High density SNP association study of a major autism linkage region on chromosome 17. *Hum Mol Genet* 2007;16:704–715. [PubMed: 17376794]
35. Bernardinelli L, Murgia SB, Bitti PP, Foco L, Ferrai R, Musu L, et al. Association between the ACCN1 gene and multiple sclerosis in Central East Sardinia. *PLoS ONE* 2007;2:e480. [PubMed: 17534430]
36. Bittel DC, Kibiryeve N, Butler MG. Whole genome microarray analysis of gene expression in subjects with fragile X syndrome. *Genet Med* 2007;9:464–472. [PubMed: 17666893]
37. Van Roy N, Vandesompele J, Berx G, Staes K, Van Gele M, De Smet E, et al. Localization of the 17q breakpoint of a constitutional 1;17 translocation in a patient with neuroblastoma within a 25-kb segment located between the ACCN1 and TLK2 genes and near the distal breakpoints of two microdeletions in neurofibromatosis type 1 patients. *Genes Chromosomes Cancer* 2002;35:113–120. [PubMed: 12203774]
38. Vandepoele K, Andries V, Van Roy N, Staes K, Vandesompele J, Laureys G, et al. A constitutional translocation t(1;17)(p36.2;q11.2) in a neuroblastoma patient disrupts the human NBP1 and ACCN1 genes. *PLoS ONE* 2008;3:e2207. [PubMed: 18493581]
39. Buyse IM, Takahashi EI, Huang S. Physical mapping of the retinoblastoma interacting zinc finger gene RIZ to D1S228 on chromosome 1p36. *Genomics* 1996;34:119–121. [PubMed: 8661032]
40. Huang S. The retinoblastoma protein-interacting zinc finger gene RIZ in 1p36-linked cancers. *Front Biosci* 1999;4:D528–532. [PubMed: 10369808]
41. Muraosa Y, Takahashi K, Yoshizawa M, Shibahara S. cDNA cloning of a novel protein containing two zinc-finger domains that may function as a transcription factor for the human heme-oxygenase-1 gene. *Eur J Biochem* 1996;235:471–479. [PubMed: 8654390]
42. Ueki N, Kondo M, Seki N, Yano K, Oda T, Masuho Y, et al. NOLP: identification of a novel human nucleolar protein and determination of sequence requirements for its nucleolar localization. *Biochem Biophys Res Commun* 1998;252:97–102. [PubMed: 9813152]
43. Wang SS, Smiraglia DJ, Wu YZ, Ghosh S, Rader JS, Cho KR, et al. Identification of novel methylation markers in cervical cancer using restriction landmark genomic scanning. *Cancer Res* 2008;68:2489–2497. [PubMed: 18381458]

44. Ikeda M, Nomura M. cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/PAS protein (BMAL1) and identification of alternatively spliced variants with alternative translation initiation site usage. *Biochem Biophys Res Commun* 1997;233:258–264. [PubMed: 9144434]
45. Bunger MK, Wilsbacher LD, Moran SM, Clendenin C, Radcliffe LA, Hogenesch JB, et al. Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell* 2000;103:1009–1017. [PubMed: 11163178]
46. Nievergelt CM, Kripke DF, Barrett TB, Burg E, Remick RA, Sadovnick AD, et al. Suggestive evidence for association of the circadian genes PERIOD3 and ARNTL with bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet* 2006;141B:234–241. [PubMed: 16528748]
47. Shi J, Wittke-Thompson JK, Badner JA, Hattori E, Potash JB, Willour VL, et al. Clock genes may influence bipolar disorder susceptibility and dysfunctional circadian rhythm. *Am J Med Genet B Neuropsychiatr Genet* 2008;147B:1047–1055. [PubMed: 18228528]
48. Partonen T, Treutlein J, Alpmann A, Frank J, Johansson C, Depner M, et al. Three circadian clock genes *Per2*, *Arntl*, and *Npas2* contribute to winter depression. *Ann Med* 2007;39:229–238. [PubMed: 17457720]
49. Mansour HA, Wood J, Logue T, Chowdari KV, Dayal M, Kupfer DJ, et al. Association study of eight circadian genes with bipolar I disorder, schizoaffective disorder and schizophrenia. *Genes Brain Behav* 2006;5:150–157. [PubMed: 16507006]
50. Nagase T, Ishikawa K, Suyama M, Kikuno R, Miyajima N, Tanaka A, et al. Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res* 1998;5:277–286. [PubMed: 9872452]
51. Missler M, Zhang W, Rohlmann A, Kattenstroth G, Hammer RE, Gottmann K, et al. Alpha-neurexins couple Ca²⁺ channels to synaptic vesicle exocytosis. *Nature* 2003;423:939–948. [PubMed: 12827191]
52. Lachman HM, Fann CS, Bartzis M, Evgrafov OV, Rosenthal RN, Nunes EV, et al. Genomewide suggestive linkage of opioid dependence to chromosome 14q. *Hum Mol Genet* 2007;16:1327–1334. [PubMed: 17409192]
53. Rodd ZA, Kimpel MW, Edenberg HJ, Bell RL, Strother WN, McClintick JN, et al. Differential gene expression in the nucleus accumbens with ethanol self-administration in inbred alcohol-preferring rats. *Pharmacol Biochem Behav* 2008;89:481–498. [PubMed: 18405950]
54. Hishimoto A, Liu QR, Drgon T, Pletnikova O, Walther D, Zhu XG, et al. Neurexin 3 polymorphisms are associated with alcohol dependence and altered expression of specific isoforms. *Hum Mol Genet* 2007;16:2880–2891. [PubMed: 17804423]
55. Kelai S, Maussion G, Noble F, Boni C, Ramoz N, Moalic JM, et al. *Nrxn3* upregulation in the globus pallidus of mice developing cocaine addiction. *Neuroreport* 2008;19:751–755. [PubMed: 18418251]
56. Young EA, Kornstein SG, Marcus SM, Harvey AT, Warden D, Wisniewski SR, et al. Sex differences in response to citalopram: A STAR *D report. *J Psychiatr Res* 2008;25:25.

Table 1

Top 25 SNP association results for response phenotype. SNPs with cluster plots that do not have distinct genotypic groups in the raw data are indicated by an * near the SNP ID.

SNP ID	rank	P value	MAF of Non-responders	MAF of Responders	OR	L95	U95	CHR	position	ancestral allele	type	closest gene	distance to gene
rs6966038	1	4.650E-07	0.19	0.25	1.64	1.35	1.99	7	156573159	A	INTERGENIC	UBE3C	-51266
rs6127921	2	3.450E-06	0.18	0.12	0.61	0.49	0.75	20	55072157	A	INTERGENIC	BMP7	106805
rs809736	3	8.190E-06	0.17	0.25	1.52	1.27	1.83	15	59117080	A	INTRONIC	RORA	0
rs43978	4	1.620E-05	0.40	0.47	1.40	1.20	1.63	17	28701273	C	INTRONIC	ACCN1	0
rs20137	5	2.060E-05	0.42	0.32	0.70	0.60	0.83	12	97786808	G	INTRONIC	ANKS1B	0
rs12883508	6	2.210E-05	0.08	0.04	0.49	0.36	0.68	14	41901648	T	INTERGENIC	LRFN5	
rs1112863	7	2.290E-05	0.08	0.04	0.48	0.34	0.67	3	12914829	T	3PRIME_UTR	IQSEC1	0
rs1858355	8	3.050E-05	0.18	0.13	0.64	0.52	0.79	3	62634698	N/A	INTRONIC	CADPS	0
rs1437956	9	3.090E-05	0.29	0.36	1.40	1.20	1.64	2	150490976	T	INTERGENIC		
rs4017754	10	3.600E-05	0.25	0.33	1.42	1.20	1.68	3	133960923	G	INTERGENIC	NPHP3;ACAD11	-36930
rs4654752	11	4.060E-05	0.38	0.46	1.37	1.18	1.59	1	227261716	A	INTERGENIC		
rs6046805*	12	4.400E-05	0.37	0.46	1.39	1.19	1.62	20	20272341	G	INTRONIC	C20orf26	0
rs2697952	13	4.970E-05	0.27	0.21	0.70	0.59	0.83	1	13998083	C	INTRONIC	PRDM2	0
rs2744692	14	5.110E-05	0.26	0.21	0.69	0.58	0.83	1	13997814	C	INTRONIC	PRDM2	0
rs17172758	15	5.730E-05	0.42	0.51	1.37	1.18	1.60	7	46646583	C	INTERGENIC		
rs9879244	16	5.740E-05	0.28	0.35	1.40	1.19	1.65	3	133971444	C	INTERGENIC	NPHP3;ACAD11	-47451
rs7239368*	17	6.120E-05	0.42	0.36	0.73	0.63	0.85	18	29899377	G	INTRONIC	NOL4	0
rs11701162	18	6.160E-05	0.24	0.31	1.41	1.19	1.66	21	42899212	C	INTERGENIC	SLC37A1	24593
rs6817919	19	6.170E-05	0.06	0.09	1.88	1.38	2.56	4	3552449	T	INTRONIC		
rs788725	20	6.660E-05	0.53	0.45	0.74	0.64	0.86	7	46121021	N/A	INTERGENIC	IGFBP3	193997
rs2368416	21	7.020E-05	0.51	0.43	0.74	0.63	0.86	10	29736991	C	INTERGENIC	SVIL	49292
rs10499638	22	7.030E-05	0.18	0.24	1.47	1.22	1.78	7	46133456	A	INTERGENIC		
rs4146387	23	7.040E-05	0.06	0.02	0.44	0.30	0.66	11	13263224	G	INTRONIC	ARNTL	0
rs17151283	24	7.380E-05	0.13	0.08	0.61	0.48	0.78	5	123538098	C	INTERGENIC		
rs10518306	25	7.690E-05	0.12	0.08	0.60	0.47	0.77	4	124955655	C	INTERGENIC		

Abbreviations: SNP ID: dbSNP128 rs identification number; MAF: minor allele frequency; OR: odds ratio; U95 and L95: upper and lower 95% confidence intervals for the OR; chr: chromosome; position: hg18 build coordinate on human genome. Regions without annotated genes within 200kb are left blank.

Table 2

Top 25 SNP association results for remission phenotype. SNPs with cluster plots that do not have distinct genotypic groups in the raw data are indicated by an * near the SNP ID.

SNP ID	rank	P value	MAF of Non-responders	MAF of Remitters	OR	L95	U95	CHR	Position	ancestral allele	type	closest gene	distance to gene
rs6966038	1	3.63E-07	0.20	0.25	1.68	1.37	2.04	7	156573159	A	INTERGENIC	UBE3C	-51266
rs6127921	2	1.07E-06	0.19	0.13	0.57	0.46	0.72	20	55072157	A	INTERGENIC	BMP7	106805
rs2830840	3	5.48E-06	0.46	0.36	0.68	0.58	0.81	21	27587218	G	INTERGENIC	EIF4A1P	73986
rs7239368	4	9.05E-06	0.42	0.36	0.69	0.59	0.82	18	29899377	G	INTRONIC	NOL4	0
rs4472078	5	1.44E-05	0.46	0.36	0.71	0.61	0.83	3	122960324	C	INTERGENIC	GOLGB1	-9032
rs6126982	6	2.98E-05	0.51	0.44	0.72	0.62	0.84	20	51889852	G	INTERGENIC	SUMO1P1	34596
rs6817919	7	3.30E-05	0.06	0.09	1.96	1.43	2.69	4	3552449	T	INTRONIC	FLJ35424	0
rs8192100	8	3.60E-05	0.15	0.11	0.61	0.49	0.77	4	140836433	A	INTRONIC	MGST2	0
rs4654712	9	3.87E-05	0.38	0.46	1.39	1.19	1.63	1	227261716	A	INTERGENIC		
rs11617613	10	4.17E-05	0.22	0.17	0.67	0.55	0.81	13	30113875	G	INTRONIC	USPL1	0
rs9864907	11	4.34E-05	0.33	0.25	0.69	0.58	0.82	3	122853753	C	INTRONIC	HCLS1	0
rs11701162	12	5.00E-05	0.24	0.30	1.43	1.20	1.71	21	42899212	C	INTERGENIC	SLC37A1	24593
rs1858385	13	5.13E-05	0.17	0.13	0.63	0.51	0.79	3	62634698	N/A	INTRONIC	CADPS	0
rs6046805	14	5.35E-05	0.38	0.46	1.40	1.19	1.64	20	20272341	G	INTRONIC	C20orf26	0
rs2914161	15	5.44E-05	0.42	0.50	1.40	1.19	1.65	5	113051338	C	INTERGENIC	YTHDC2	92460
rs4146387	16	5.75E-05	0.05	0.02	0.41	0.26	0.63	11	13263224	G	INTRONIC	ARNTL	0
Rs439783	17	5.87E-05	0.41	0.46	1.39	1.18	1.62	17	28701273	C	INTRONIC	ACCN1	0
rs2989579	18	6.04E-05	0.26	0.21	0.69	0.57	0.82	9	80175562	C	INTERGENIC	PSAT1;C8orf62	40735
rs11128623	19	6.14E-05	0.08	0.04	0.47	0.33	0.68	3	12914829	T	3PRIME_UTR	IQSEC1	0
rs10512091	20	6.39E-05	0.17	0.23	1.51	1.23	1.84	9	81612652	G	INTERGENIC	TLE4	-81427
rs1468864	21	6.63E-05	0.01	0.02	3.86	1.99	7.49	7	104078689	A	INTRONIC	LHFPL3	0
rs2744692	22	7.11E-05	0.25	0.21	0.69	0.57	0.83	1	13997814	C	INTRONIC	PRDM2	0
rs5999086	23	7.19E-05	0.24	0.17	0.66	0.54	0.81	22	32476368	G	INTRONIC	LARGE	0
rs4257844	24	7.42E-05	0.16	0.20	1.57	1.25	1.95	6	67692405	T	INTERGENIC		
rs809736	25	7.64E-05	0.17	0.24	1.47	1.22	1.79	15	59117080	A	INTRONIC	RORA	0

Abbreviations: SNP ID: dbSNP128 rs identification number; MAF: minor allele frequency; OR: odds ratio; U95 and L95: upper and lower 95% confidence intervals for the OR; chr: chromosome; position: hg18 build coordinate on human genome. Regions without annotated genes within 200kb are left blank.